The Dissertation Committee for Zhihua Li Certifies that this is the approved version of the following dissertation:

From Developing Protein-Protein Interaction Strategies to Identifying Gene Functions – Case Studies for Transcription Factor Complexes and Ribosome Biogenesis Genes

Committee:

Edward M. Marcotte, Supervisor

Dean R. Appling

David W. Hoffman

Vishwanath R. Iyer

Scott W. Stevens

From Developing Protein-Protein Interaction Strategies to Identifying Gene Functions – Case Studies for Transcription Factor Complexes and Ribosome Biogenesis Genes

by

Zhihua Li, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Dedication

This dissertation is dedicated to my parents for their love and encouragement throughout my life.

Acknowledgements

I would like to thank my supervisor, Dr. Edward Marcotte, for his patience, encouragement, and support throughout my graduate career. I also would like to acknowledge Dr. Arlen Johnson for his great support and guidance in the ribosome biogenesis project. I am grateful to Dr. Vishwanath Iyer for his support in the DNA microarray experiments. I also thank Dr. David Hoffman for help with NMR data collection, and Dr. Scott Stevens for his useful suggestions on affinity purification and ribosome biogenesis. I am also grateful to Dr. Dean Appling for his suggestions on the carbon and nitrogen metabolism, and his proof-reading and comments on this dissertation. I want to thank Marcotte lab former and current members for their friendship and help, especially John Prince for his generous help with mass spectrometry analysis, Emily Moradi for her microscopy image analysis, and Dr. Insuk Lee and Kris McGary for their gene network analysis. I also thank Iyer lab members for their help with DNA microarray experiments, especially Dr. Zhanzhi Hu and Dr. Jian Hu for their useful discussions. I thank Dr. Nai-Jung Hung for help with ribosome export experiments, and Kai-Yin Lo and Reed for help with mRNA export assays.

From Developing Protein-Protein Interaction Strategies to Identifying Gene Finctions – Case Studies for Transcription Factor Complexes and Ribosome Biogenesis Genes

Publication No._____

Zhihua Li, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Edward M. Marcotte

Protein-protein interactions are central to their biological functions in cells. Many approaches have been applied to study protein-protein interactions in a genomic-scale. In an attempt to develop new strategies to study protein-protein interactions, FRET by using ECFP and EYFP as the donor and receptor was evaluated for possible application in protein-protein interaction study in a high-throughput fashion. Due to the intrinsic properties of ECFP and EYFP, FRET-based protein-protein interaction assay is not suitable for large-scale studies. Instead, tandem affinity purification coupled with mass spectrometry approach proved to be a useful strategy to identify protein interacting partners. Several transcription factor complexes in yeast were successfully purified and novel components in the complexes were identified by combining a shotgun mass spectrometry approach and a differential analysis of the mass spectrometry data. In particular, a negative regulator of G1 to S phase transition during cell cycle, Whi5p, was identified to be a component of SBF complex; a regulator of nitrogen metabolism, Gln3p,

was identified to be a component of Hap2/3/5 complex that regulates carbon metabolism, suggesting a crosstalk between nitrogen and carbon metabolism. Additionally, one-step purification coupled with shotgun mass spectrometry analysis was applied to simplify and improve the affinity purification approach used for protein-protein interaction studies. In order to map protein complexes in their native state, a sucrose density gradient was used to separate protein complexes in cells. The proteins within each fraction from the sucrose density gradient were analyzed and quantified with mass spectrometry to obtain the protein abundance profiles across the gradient. The known protein complexes were identified by clustering the protein abundance profiles. This method could possibly be improved to become a generic approach to mapping protein complexes. The goal of protein-protein interaction studies is to determine the protein functions. In an effort to identify ribosome biogenesis genes from a yeast gene network reconstructed from diverse large-scale interaction data sets, at least 25 new ribosome biogenesis genes were confirmed by extensive experimental validations, underscoring the value of protein-protein interaction studies and gene interaction network.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: General Introduction

Genome sequencing has generated over 1,000 complete whole genome sequences (http://wit.integratedgenomics.com/GOLD) and the number continues to increase. However, functional annotations of the genes encoded in the genomes lag behind. The demand of interpreting the whole genome leads to the development of systematic analyses for the complete set of major molecules in cells: RNAs (transcriptome), proteins (proteome), and metabolites (metabolome), which has been termed functional genomics (Oliver, 2002). Additionally, all the molecules in cells connect and cooperate to each other to form a system to eventually function in cellular behavior (Illustration 1.1A). Therefore, it becomes increasingly important to understand the cells at the system level, requiring complete description of interactions among different molecules and dynamic natures of the interactions. Among the interactions, protein-protein interactions draw increasing attention not only because proteins are central to biological function but also because many proteins are candidates for drug developments.

A wealth of methods have been developed to detect protein-protein interactions (Fu, 2004; Golemis and Adams, 2005; Shoemaker and Panchenko, 2007a; Shoemaker and Panchenko, 2007b), many of which are suitable for functional genomics studies and have been successfully applied to high-throughput identifications of protein-protein interactions (Illustration 1.1B).

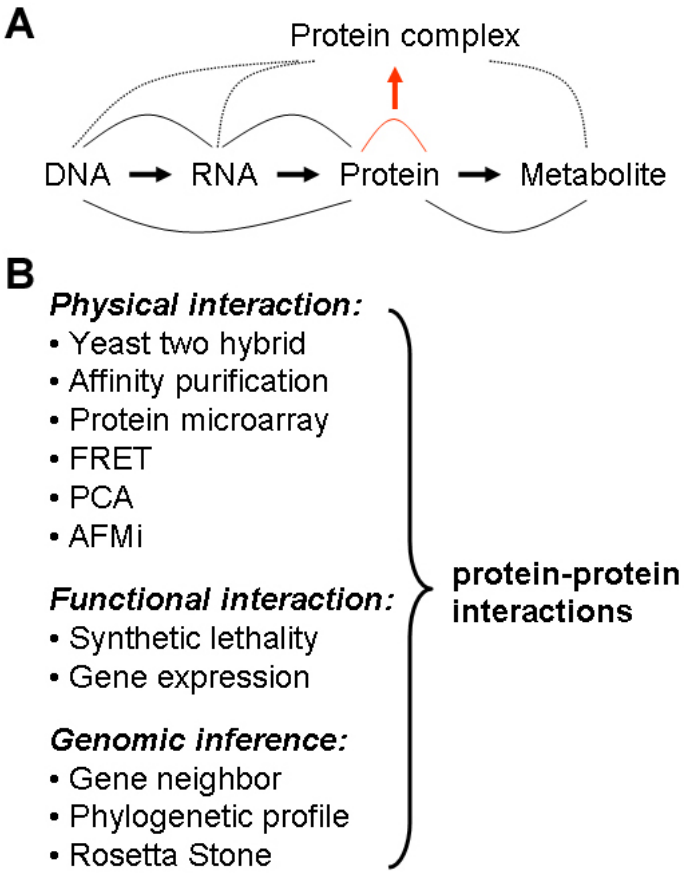


Illustration 1.1 Major molecules in cells and their interactions to each other.
(A) A diagram of modified central dogma and interactions among major molecules in cells (solid lines). Protein-protein interactions (red solid line) form protein complexes that can further interact with other molecules (dashed lines).
(B) Major methods used to detect protein-protein interactions. Fluorescence resonance energy transfer (FRET), protein complementation assay (PCA), and atomic force microscopy imaging (AFMi) methods have not been applied in a high-throughput fashion, and are possible options in the future.

Yeast two-hybrid method (Y2H) has been widely used for genome-wide interaction studies in different organisms, such as yeast (Ito et al., 2001; Uetz et al., 2000), worm (Li et al., 2004), fly (Giot et al., 2003), and human (Ghavidel et al., 2005; Rual et al., 2005). It is based on a transcription assay in which transcription occurs only if a protein that is fused with the activation domain of the transcription activator interacts with another protein that is fused with the DNA binding domain of the transcription activator (Fields and Song, 1989). Y2H is an *in vivo* assay that detects protein binary physical interactions. However, this assay is known to produce high false positives and false negatives, attributed by nonspecific interactions and protein fusions that could potentially disrupt interactions (Deeds et al., 2006).

Another method is based on tandem affinity purification (TAP) coupled with mass spectrometry that allows purification of protein complexes under mild conditions and reducing nonspecific interactions (Rigaut et al., 1999). This method was successfully applied in large-scale mapping protein complexes in yeast under physiological conditions because each gene can be tagged at its genomic locus (Gavin et al., 2006; Gavin et al., 2002; Krogan et al., 2006). Similar approach was also applied to identify E. coli protein complexes (Butland et al., 2005). Affinity purification under physiological conditions has the advantage of retaining proteins' stoichiometries within the complex and identifying protein associations that otherwise can not be detected by binary interaction detection method. However, false positives still exist due to nonspecific interactions during purification and it will be worse if the bait protein is over-expressed in cells (Ho et al., 2002). In human cells, the genes can not be easily tagged at its genomic locus, thereby limiting its utility. Nevertheless, the affinity purification method was recently applied to map protein complexes in a large-scale fashion by expressing a bait protein fused with a single tag in human cells (Ewing et al., 2007).

Protein microarray is another useful technology to study protein-protein interactions and enzyme activities (MacBeath and Schreiber, 2000; Zhu et al., 2001). Additionally, protein microarray can be utilized to quantitatively measure the dissociation constant of protein interactions with accuracy comparable to surface plasmon resonance (SPR) technology (Jones et al., 2006). Currently protein microarrays are mainly used to study interactions for target proteomes, and screening protein interactions for every protein in the cell by using protein microarray would be a daunting task. It is noteworthy that label-free detection methods for protein microarray have been developed. These methods take the advantage of recent development in high-throughput SPR imaging and atomic force microscopy (AFM) technologies to simultaneously analyze thousands protein interactions (Boozer et al., 2006; Yu et al., 2006).

Instead of detecting protein physical associations, gene coexpression method and synthetic lethality method can be utilized to assess protein functional associations, from which the relationships among genes and pathways can be derived. Gene coexpression method by using DNA microarray technology is high-throughput, genomic scale, and widely applicable for any organism. Proteins with similar functions tend to co-express (Eisen et al., 1998) and direct interacting proteins are more likely to co-express compared with non-interacting proteins (Jansen et al., 2002). Synthetic lethality method has been widely and successfully used in yeast genetic studies. The development of synthetic genetic arrays (SGA) provided a genome-wide genetic interaction map (Tong et al., 2001; Tong et al., 2004). The synthetic lethality method could also be applied to higher organisms such as C. elegans by using RNAi technology to inhibit gene functions (Lehner et al., 2006).

Genomic context information can also be utilized to infer gene/protein functional associations. Methods using genomic context information include gene neighbor or gene

cluster method, phylogenetic profile method, and Rosetta Stone method. Co-regulated genes tend to cluster together in the genome and perform similar functions in the cell. By analyzing such clusters or operons , the functional linkages among genes can be derived (Overbeek et al., 1999). Phylogenetic profile method is based on the hypothesis that functionally linked proteins co-evolve and tend to exist together across different organisms (Pellegrini et al., 1999). Rosetta Stone method is based on the observation that two interacting proteins have orthologs in other genomes that are fused into one protein (Enright et al., 1999; Marcotte et al., 1999a). All these methods rely on the available genome sequences and computational analyses that are relatively cheaper and easier than experimental methods. However, each method addresses a fraction of interactions among genes, and many of those computationally predicted interactions have not been confirmed by experimental methods.

Some methods used for detecting protein-protein interactions have not been applied in a high-throughput fashion but hold promise in the future. Fluorescence resonance energy transfer (FRET) is one such method based on the energy transfer from a donor molecule to an acceptor molecule. FRET signal is observed when one protein labeled with donor fluorophore interacts with another protein labeled with acceptor fluorophore to bring the two fluorophores within a short distance (less than 100 Å) (Stryer, 1978). When proteins are fused to fluorescent proteins, FRET can be used for *in vivo* studies (Miyawaki et al., 1997). It measures direct physical interactions and can be applied to measure the dynamic interactions between proteins (Ting et al., 2001). Another promising method for high-throughput study is protein-fragment complementation assay (PCA). Two proteins are fused to complementary fragments of a reporter protein. If these two proteins interact to each other, the reporter fragments are brought together to reconstitute the reporter activity that can be monitored as a readout for protein-protein

interactions (Michnick et al., 2000). The assay based on dihydrofolate reductase (DHFR) activity is especially suitable for large-scale studies because of the simple survival selection (Remy and Michnick, 1999).

Different methods for protein-protein interaction studies have their own merits and limitations. Accordingly, the data generated from different methods have their underlying bias and the data quality varies considerably in terms of coverage and accuracy (von Mering et al., 2002). Data evaluation and integration is an ideal approach to reduce the bias, increase the coverage, and improve the accuracy. The integrated protein-protein interactions can therefore be utilized to reconstruct an interaction network for the proteome (Jansen et al., 2003; Lee et al., 2004; Lee et al., 2007). The resulting network is a rich resource that can be used to predict and elucidate protein functions (Sharan et al., 2007).

In this dissertation, I will first introduce some of my attempts to develop protein-protein interaction strategies: In chapter two, FRET method will be evaluated for the feasibility to be utilized in large-scale protein-protein interaction studies; In chapter three, affinity purification coupled with mass spectrometry method will be utilized to purify several transcription factor complexes, from which new gene functions can be inferred; In chapter four, a new method based on fractionation of cell lysate and mass spectrometry identification to map protein interactions will be discussed. Then in chapter five, I will describe a large-scale experimental study to identify new ribosome biogenesis genes from a yeast functional gene network reconstructed from existing large-scale heterogeneous interaction data.

Chapter 2: Detection of Protein-Protein Interactions by FRET

2.1 INTRODUCTION

FRET describes the energy transfer from a donor molecule to an acceptor molecule. The energy transfer theory is first described by Foster and further experimentally proved in well-defined model systems (Stryer, 1978). The energy transfer efficiency is mathematically defined as $E = 1/(1 + (r/R_0)^6)$, where r is the distance between the centers of the donor and acceptor fluorophores and R_0 is the distance at which the transfer efficiency is 50%. Due to the strong dependence on the distance between the donor and the acceptor, FRET was used as a spectroscopic ruler that could measure a distance at 10-60 Å range (Stryer and Haugland, 1967). The energy transfer efficiency also depends on the angular relationship between the donor and acceptor, the overlap of the emission spectrum of the donor with the excitation spectrum of the acceptor, the rate constant of fluorescence emission by the donor, and the quantum yield of fluorescence of the donor in the absence of the acceptor (Stryer, 1978). Any donor and acceptor fluorescence molecules that satisfy those conditions could be used for FRET.

For protein-protein interaction studies, fluorescence proteins are favorable because they can be encoded genetically. A blue variant of green fluorescence protein (GFP) (BFP5) and a red-shifted variant of GFP (RSGFP4) were first used for FRET assay to monitor the Factor Xa protease activity (Mitra et al., 1996). Then enhanced cyan and yellow fluorescence proteins (ECFP and EYFP) were utilized for FRET to measure the dynamics of Ca^{2+} signals in a single cell, and this fluorescence protein pair improved the brightness and signal-to-noise ratio for FRET, when compared to BFP and GFP as the FRET pair (Miyawaki et al., 1997).

I started to evaluate the feasibility of a large-scale FRET assay by using ECFP and EYFP as a FRET pair. First, a fusion protein of ECFP and EYFP as FRET positive

control was constructed. Then known interaction partners were individually fused with ECFP and EYFP for FRET assay. If FRET could be easily applied for interaction assays, two cDNA expression libraries could be constructed with each gene tagged with either ECFP or EYFP gene. After transforming a pair of genes into the cell, FRET signal could be monitored by either microscopy or flow cytometry.

2.2 MATERIALS AND METHODS

2.2.1 Plasmid construction

pETCFP-C construction: ECFP gene was obtained by PCR from pECFP (Clontech) by using oligos HX-F and HX-R (Table 2.1). The PCR fragment was digested with HindIII and XhoI, and ligated into expression vector pET28a(+) (Novagen). The resulting plasmid can be used to express a fusion protein with ECFP at the carboxyl-terminus.

pETYFP-C construnction: EYFP gene was obtained by PCR from pEYFP (Clontech) by using oligos HX-F and HX-R (Table 2.1). The PCR fragment was digested with HindIII and XhoI and ligated into expression vector pET28a(+) (Novagen). The resulting plasmid can be used to express a fusion protein with EYFP at the carboxyl-terminus.

pETYCFP construction: EYFP gene was obtained by PCR from pEYFP (Clontech) by using oligos NE-F and NE-R (Table 2.1). The PCR fragment was digested with NcoI and EcoRI, and ligated into the vector pETCFP-C. The resulting plasmid is used to express the EYFP and ECFP fusion protein.

pETCFP-N construction: ECFP gene was obtained by PCR from pECFP (Clontech) by using oligos NE-F and NE-R (Table 2.1). The PCR fragment was digested

with NcoI and EcoRI, and ligated into expression vector pET28a(+) (Novagen). The resulting plasmid can be used to express a fusion protein with ECFP at the amino-terminus.

pETYFP-N construction: EYFP gene was obtained by PCR from pEYFP (Clontech) by using oligos NE-F and NE-R (Table 2.1). The PCR fragment was digested with NcoI and EcoRI, and ligated into expression vector pET28a(+) (Novagen). The resulting plasmid can be used to express a fusion protein with EYFP at the amino-terminus.

pRC construction: The coding region for three WW domains of *RSP5* was amplified by PCR with oligos RSP5-WW-F and RSP5-WW-R (Table 2.1). The PCR product was digested with NdeI and EcoRI, and ligated into pETCFP-C. The resulting vector was used express a fusion protein with WW domains at the amino-terminus and ECFP at the carboxyl-terminus.

pRY construction: The coding region for three WW domains of *RSP5* was amplified by PCR with oligos RSP5-WW-F and RSP5-WW-R (Table 2.1). The PCR product was digested with NdeI and EcoRI, and ligated into pETYFP-C. The resulting vector was used express a fusion protein with WW domains at the amino-terminus and EYFP at the carboxyl-terminus.

pPC construction: The oligos PPXY-F and PPXY-R were incubated together at 95°C for 5 minutes, and then cooled down at room temperature to anneal. The annealed fragment was ligated into pETCFP-C between BamHI and EcoRI. The resulting vector was used to express a fusion protein with the PPXY motif-containing peptide at the amino-terminus and ECFP at the carboxyl-terminus.

pPY construction: Similar to pPC construction, the annealed fragment was ligated into pETYFP-C between BamHI and EcoRI. The resulting vector was used to express a

fusion protein with the PPXY motif-containing peptide at the amino-terminus and EYFP at the carboxyl-terminus.

2.2.2 Measuring fluorescence and FRET signals

For protein expression, the plasmid was first transformed into E. coli strain BL21(DE3). After culturing the cells to OD_{600} (optical density at 600 nm) 0.4 in Luria-Bertani Medium (LB medium), 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression for 2-4 hours. Cells were concentrated and fluorescence was monitored by a fluorescence microscope (Nikon) using a filter for GFP signal or a scanning spectrofluorometer (Photon Technologies International Quanta Master Model C). FRET signal from whole cells or cell extracts was monitored by fixing excitation wavelength at 410 nm to scan the emission spectrum from 420 nm to 625 nm with the scanning spectrofluorometer.

Table 2.1 Oligo sequences for vector construction

Oligo name	**Sequence**	**Vendor**
HX-F	CCCAAGCTTCTATGGTGAGCAAGGGCGAG	Life Technologies
HX-R	CCGCTCGAGCTTGTACAGCTCGTCCATG	Life Technologies
NE-F	CATGCCATGGTGAGCAAGGGCGAG	Life Technologies
NE-R	GGAATTCCTTGTACAGCTCGTCCATG	Life Technologies
RSP5-WWW-F	GCCATGGGCCATATGGCTACCAGACAATACTCTTC	Sigma
RSP5-WWW-R	CGGAATTCTAGCGATGATGGAAGTCTTGG	Sigma
PPXY-F	GATCCCCACCACCTGGTTATGGTTCTGGTTCTG	Sigma
PPXY-R	AATTCAGAACCAGAACCATAACCAGGTGGTGGG	Sigma

2.3 RESULTS

2.3.1 Expression of ECFP, EYFP, and EYFP-ECFP fusion proteins in E. coli

ECFP, EYFP, and EYFP-ECFP were expressed in E. coli from vectors pETCFP-C, pETYFP-C, pETCFP-N, pETYFP-N, and pETYCFP by IPTG induction. Protein expression was first monitored by directly detecting fluorescence using a fluorescence microscope (Figure 2.1). Fluorescence signal was not observed for cells without IPTG induction. Cells with pETYFP-C vector showed bright signal, indicating strong expression of EYFP from pETYFP-C (Figure 2.1B), whereas weak signal was observed for ECFP from pETCFP-C (Figure 2.1A). However, no fluorescence signal was observed for cells with either pETCFP-N or pETYFP-N (Figure 2.1C), suggesting that there was no protein expression or the expressed proteins were not functional. The fluorescence of EYFP-ECFP fusion protein was observed after IPTG induction (Figure 2.1D). Additionally, fluorescence signal was observed after IPTG induction at both 30°C and 37°C (Figure 2.1A, 2.1B, and 2.1D). pETCFP-N and pETYFP-N were not used for further study because fluorescence signals were not observed after IPTG induction.

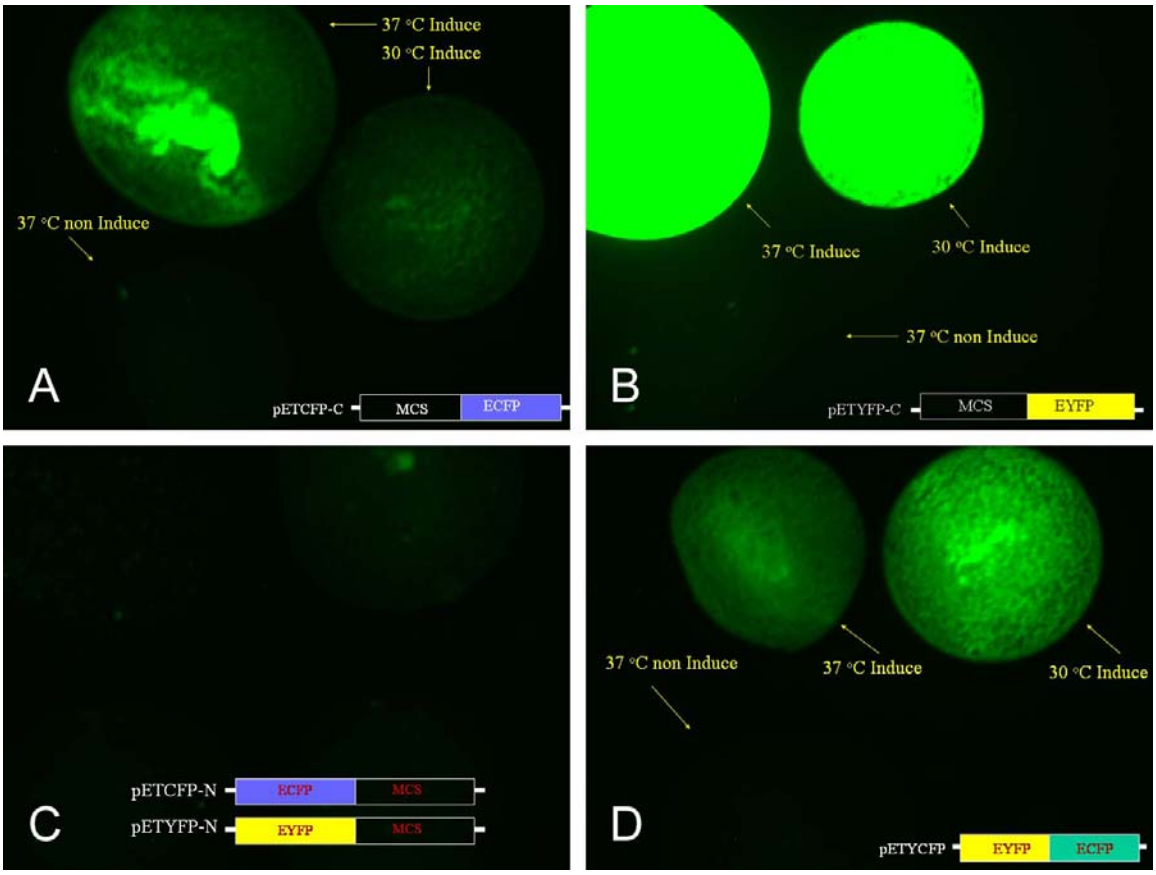


Figure 2.1 Protein expressions monitored by fluorescence microscope.
(A) Fluorescence signals monitored for concentrated E. coli cells containing pETCFP-C under induction and non-induction conditions. A schematic diagram of the vector is shown at the bottom, and MCS represents a multiple cloning site.
(B) Fluorescence signals monitored for concentrated E. coli cells containing pETYFP-C under induction and non-induction conditions.
(C) Fluorescence signals monitored for concentrated E. coli cells containing pETYFP-N and pETCFP-N under induction and non-induction conditions.
(D) Fluorescence signals monitored for concentrated E. coli cells containing pETYCFP under induction and non-induction conditions.

To further estimate the proportion of soluble proteins after IPTG induction, SDS-PAGE was used to compare soluble fraction and insoluble fraction after lysing the cells by sonication. ECFP, EYFP, and EYFP-ECFP were expressed in a significant amount after IPTG induction, with more protein expression observed after IPTG induction at 37°C (Figure 2.2). At 30°C induction, about half of ECFP was in soluble fraction, whereas about one third of ECFP was in soluble fraction after induction at 37°C (Figure 2.2A). Similarly, more EYFP was in soluble fraction after induction at 30°C when compared to induction at 37°C (Figure 2.2A). However, most of EYFP-ECFP fusion proteins existed in insoluble fraction after induction at either 30°C or 37°C (Figure 2.2B).

Protein expressions were also monitored by using a fluorometer. When the excitation wavelength was fixed at 453 nm and the emission spectrum was obtained by scanning from 463 nm to 650 nm, fluorescence signals were observed for ECFP, EYFP, and EYFP-ECFP after IPTG induction, whereas no fluorescence was observed without IPTG induction (Figure 2.3).

All the above data showed that ECFP, EYFP, and EYFP-ECFP proteins were expressed in E. coli by IPTG induction from vectors pETCFP-C, pETYFP-C, and pETYCFP, respectively, and at least a fraction of expressed proteins was functional.

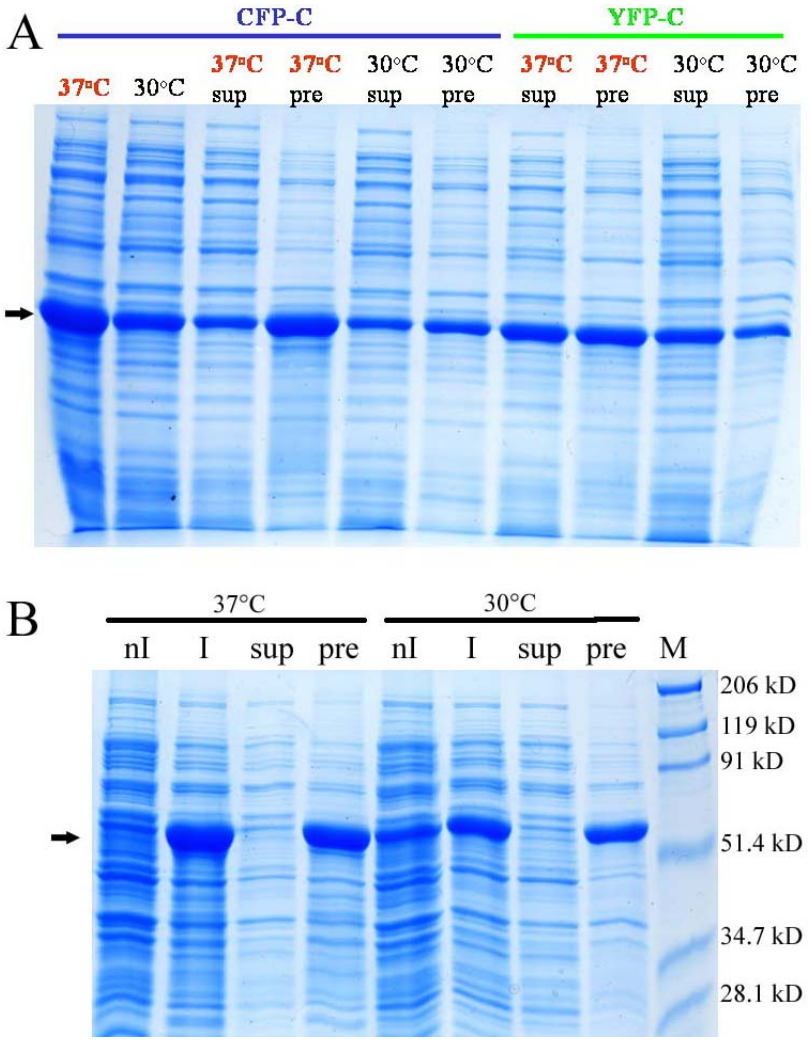


Figure 2.2 SDS-PAGE analyses of protein expressions for ECFP, EYFP, and EYFP-ECFP fusion proteins.

(A) ECFP and EYFP were expressed in E. coli at either 30°C or 37°C. Equivalent amounts of total proteins from soluble fraction (sup) and insoluble fraction (pre) were loaded onto a 12% SDS-PAGE gel. CFP-C and YFP-C represent ECFP and EYFP expressed from vectors pETCFP-C and pETYFP-C, respectively. The first two lanes for CFP-C were loaded with total lysates. The arrow indicates the expressed ECFP and EYFP with a molecular weight about 30 kD.

(B) EYFP-ECFP fusion proteins were expressed in E. coli at either 30°C or 37°C, and protein samples were loaded onto 10% SDS-PAGE gel. No expressed proteins were observed in the whole cell lysate under non-induction condition (nI), and significant amounts of proteins in the whole cell lysate were observed under induction condition (I). Most expressed proteins were in the insoluble fractions (pre), as compared to the soluble fractions (sup). Molecular markers (M) are in the last lane and labeled on the right side.

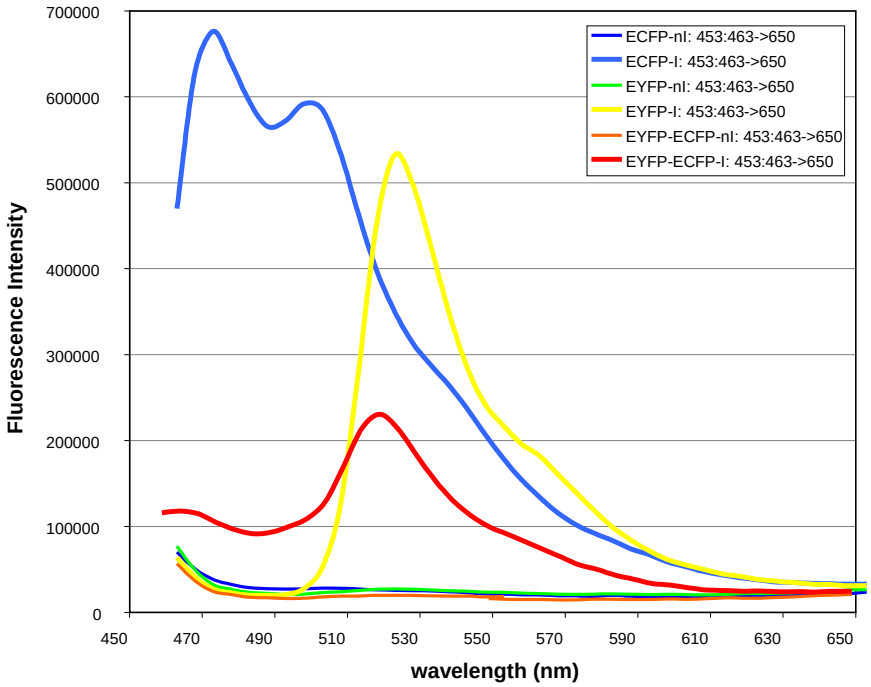


Figure 2.3 Protein expressions monitored by a spectrofluorometer.
E. coli cells with expression vectors cultured at non-induction (nI) and induction (I) conditions were used for scanning the emission spectrum from 463 nm to 650 nm by fixing the excitation wavelength 453 nm. Fluorescence emission spectra of ECFP, EYFP, and EYFP-ECFP fusion protein were not observed under IPTG non-induction conditions (blue, green, and orange colors). Fluorescence emission spectra of ECFP, EYFP, and EYFP-ECFP fusion protein were observed under IPTG induction conditions (cyan, yellow, and red colors).

2.3.2 Spectra properties of expressed ECFP and EYFP

Excitation and Emission spectra of ECFP or EYFP were obtained by a spectrofluorometer from IPTG-induced cells containing plasmids pETCFP-C or pETYFP-C (Figure 2.4). The maximum excitation wavelengths of ECFP and EYFP are 434 nm and 514 nm, respectively, and the maximum emission wavelength of ECFP and EYFP are 476 nm and 527 nm, respectively. Lower fluorescence intensity was observed for ECFP compared to the fluorescence intensity of EYFP, which was consistent with the observation of five-times brighter of EYFP than ECFP (Day et al., 2001). The emission spectrum of ECFP overlaps with the excitation spectrum of EYFP, making ECFP and EYFP possible for FRET assays.

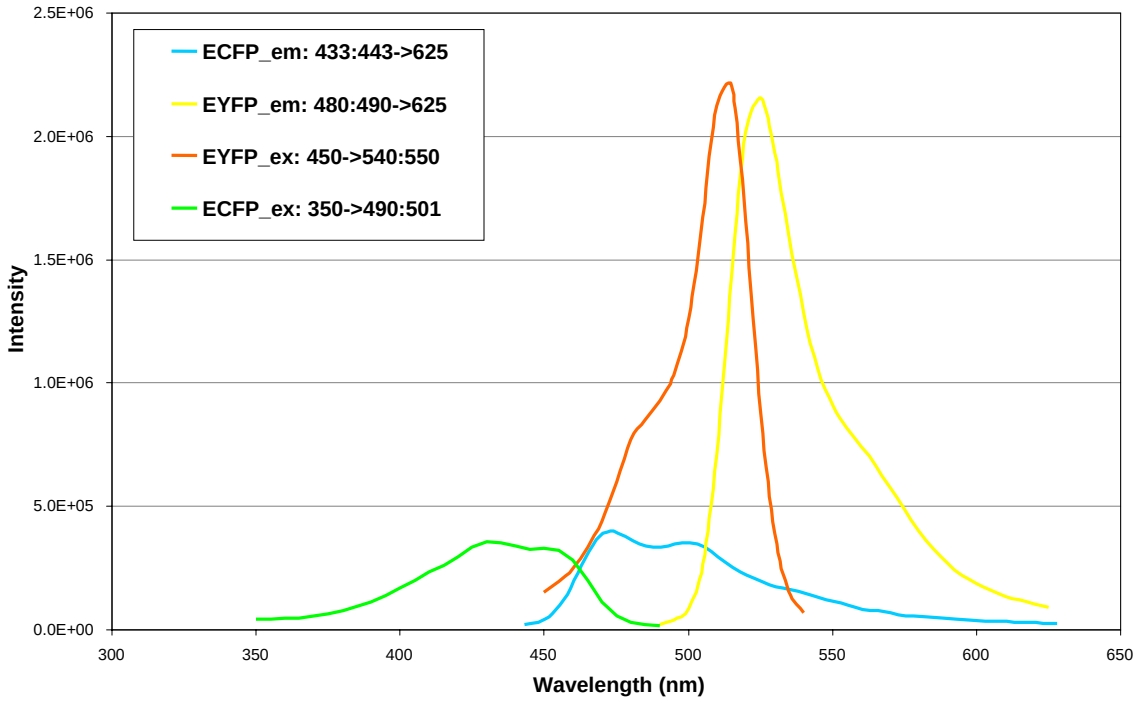


Figure 2.4 Spectra properties of ECFP and EYFP.
E. coli cells with pETCFP, pETYFP, and pETYCFP were culture under IPTG induction conditions. The emission spectrum of ECFP was acquired by scanning from 443 nm to 625 nm at a fixed excitation wavelength 433 nm (ECFP_em, cyan color), and the excitation spectrum of ECFP was obtained by scanning from 350 nm to 490 nm at a fixed emission wavelength 501 nm (ECFP_ex, green color). The emission spectrum of EYFP was acquired by scanning from 490 nm to 625 nm at a fixed excitation wavelength 480 nm (EYFP_em, yellow color), and the excitation spectrum of EYFP was obtained by scanning from 450 nm to 540 nm at a fixed emission wavelength 550 nm (EYFP_ex, red color).

2.3.3 Detecting FRET signals of EYFP-ECFP fusion proteins

To detect FRET signals for EYFP-ECFP fusion proteins, E. coli cells with expressed EYFP-ECFP fusion proteins were monitored with a spectrofluorometer by fixing the excitation wavelength at 410 nm and scanning the emission spectrum from 420 nm to 625 nm. Two peaks were observed from the emission spectrum, one at 476 nm corresponding to the maximum excitation wavelength of ECFP and another one at 527 nm corresponding to the maximum excitation wavelength of EYFP (Figure 2.5). However, EYFP itself can also be excited at 410 nm because a lower intensity peak at 527 nm was observed for EYFP compared to the peak for EYFP-ECFP fusion protein (Figure 2.5). Instead, if EYFP-ECFP fusion protein and EYFP were excited at 450 nm that is more close to the maximum excitation wavelength of EYFP, a higher emission peak of EYFP was observed when compared to the emission peak of EYFP-ECFP fusion protein, indicating that more functional EYFP proteins in the cells expressing EYFP alone than that in the cells expressing EYFP-ECFP fusion proteins. This means that the peak at 527 nm observed for EYFP-ECFP at excitation wavelength 410 nm is a mixture of EYFP emission signals from EYFP itself and FRET signals due to energy transfer from ECFP emission. One possible solution to reduce the emission signals from EYFP itself is to lower the excitation wavelength to less than 410 nm. However, this will also reduce the excitation of ECFP, therefore reducing the FRET signals. Nonetheless, FRET signals were observed from EYFP-ECFP fusion proteins.

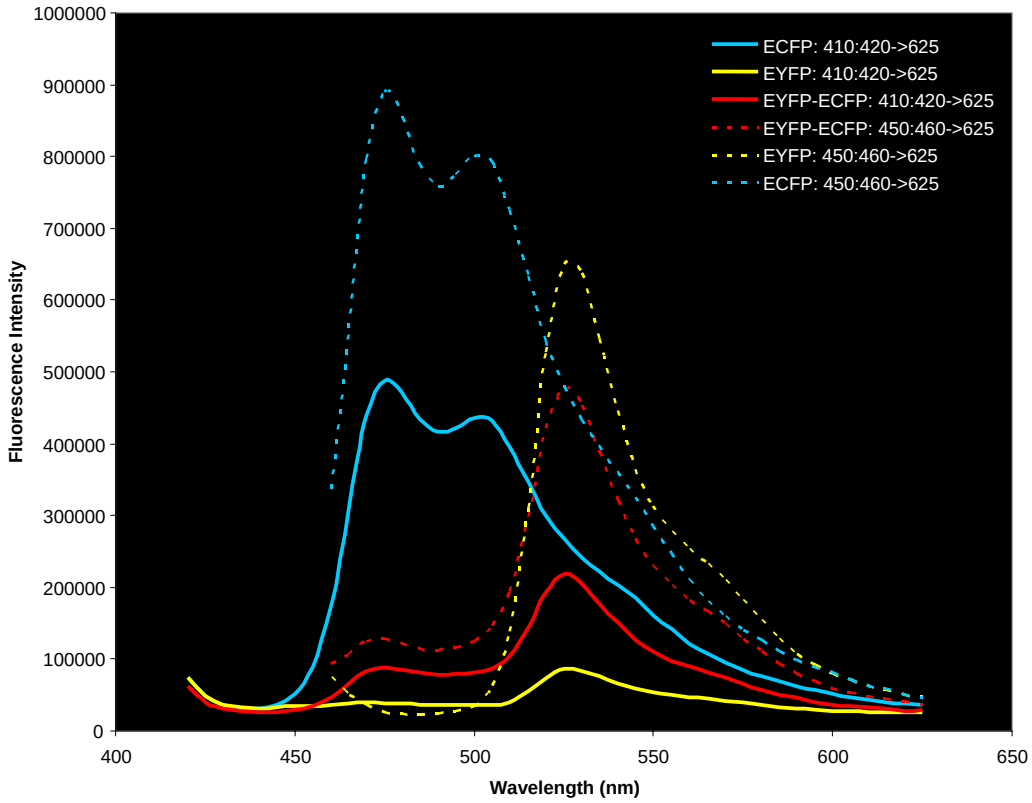


Figure 2.5 FRET of EYFP-ECFP fusion protein.
E. coli cells with expressed EYFP-ECFP fusion proteins were excited at either 410 nm or 450 nm, and FRET signals were observed by scanning from 420 nm to 625 nm (solid red line and dashed red line for excitation at 410 nm and 450 nm, respectively). Cells with expressed EYFP were used as a control to monitor the emission signals from EYFP itself (solid yellow line and dashed yellow line for excitation at 410 nm and 450 nm, respectively). Cells with expressed ECFP were also monitored at excitation wavelength 410 nm (solid cyan line) and 450 nm (dashed cyan line).

2.3.4 Expression of WW domains and PPXY-containing peptides fused with either ECFP or EYFP

Four fusion proteins were expressed in E. coli and observed under a fluorescence microscope (Figure 2.6). EYFP or ECFP was fused to carboxyl-terminus of a peptide that contains a PPXY motif to form PY or PC that showed fluorescence under the microscope after IPTG induction (Figure 2.6A), indicating functional expression of PY or PC. However, there was also significant fluorescence without IPTG induction, especially for PY, possibly due to the expression leaking from the promoter. Similar to EYFP and ECFP expression, EYFP fusion protein PY gave much stronger florescence signals that ECFP fusion protein PC (Figure 2.6A). EYFP was fused to carboxyl-terminus of WW domain of Rsp5p to form RY that showed weak fluorescence under the microscope after IPTG induction (Figure 2.6B); ECFP was fused to carboxyl-terminus of WW domain of Rsp5p to form RC that showed very weak fluorescence after IPTG induction (Figure 2.6C). The low fluorescence for RY and RC are possibly due to low expression of the fusion proteins or low functional proteins in the cells.

To examine the total expression levels and the soluble protein expression levels of PY, PC, RY, and RC, SDS-PAGE gels were applied. PY and PC were expressed in large amounts after IPTG induction, and about one third of the total expressed proteins were in soluble fraction (Figure 2.7A). Additionally, proteins were also expressed in cells without IPTG induction, consistent with the observation of fluorescence under the microscope (Figure 2.7A). RY and RC showed very low expression after IPTG induction (Figure 2.7B), consistent with weak fluorescence signals observed under the microscope. Low expression of RY or RC might be attributed to WW domain that was fused to the amino-terminus of EYFP or ECFP.

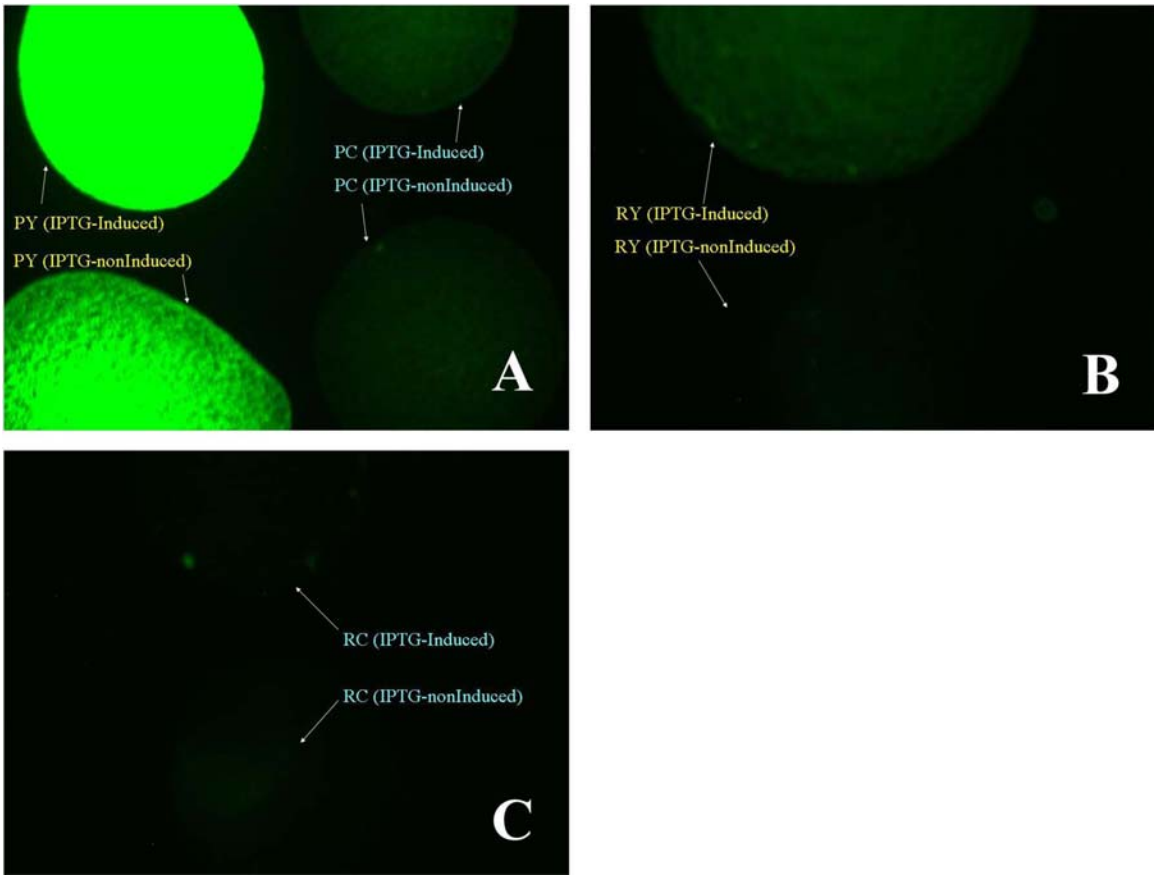


Figure 2.6 Expression of fusion proteins (PY, PC, RY, and RC) monitored by fluorescence microscopy.
(A) E. coli cells with expressed fusion protein PY (PPXY motif-containing peptide fused at the amino-terminus of EYFP) or PC (PPXY motif-containing peptide fused at the amino-terminus of ECFP) were monitored by using the GFP filter in a fluorescence microscope under IPTG induced and non-induced conditions.
(B) Cells with expressed fusion protein RY (WW domain from Rsp5p fused at the amino-terminus of EYFP) were monitored under IPTG induced and non-induced conditions.
(C) Cells with expressed fusion protein RC (WW domain from Rsp5p fused at the amino-terminus of ECFP) were monitored under IPTG induced and non-induced conditions.

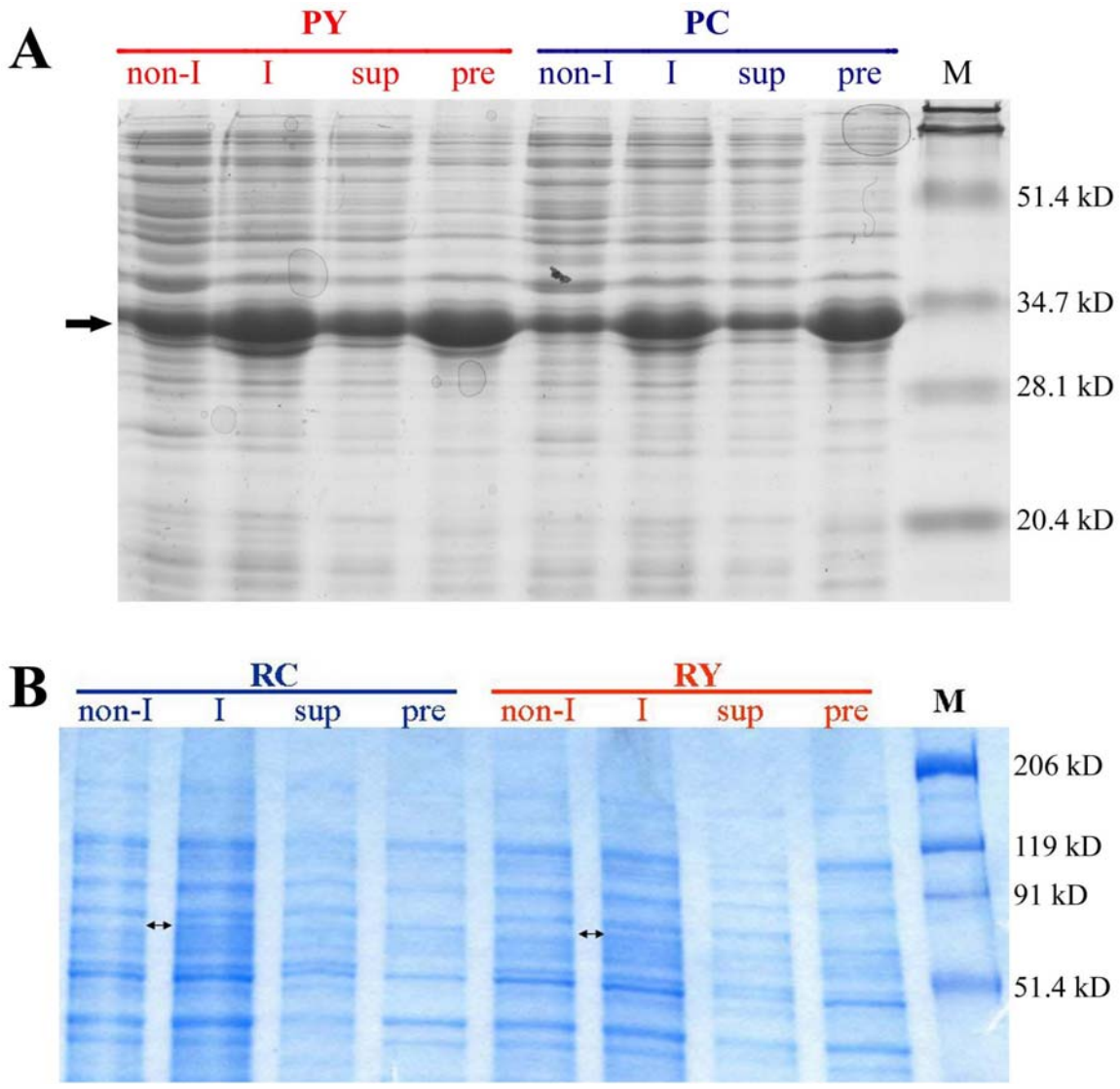


Figure 2.7 Expression of fusion proteins (PY, PC, RY, and RC) monitored by SDS-PAGE.
(A) E. coli cells with expression vector for fusion proteins PY or PC were either induced or not induced by IPTG. Total cell lysates from non-induced sample (non-I) and induced sample (I), and the soluble fraction (sup) and insoluble fraction (pre) of the IPTG-induced samples were loaded onto 12% SDS-PAGE. The expressed fusion proteins are indicated by the arrow, and the molecular weight markers are labeled on the right side.
(B) E. coli cells with expression vector for fusion proteins RC or RY were either induced or not induced by IPTG. Total cell lysates from non-induced sample (non-I) and induced sample (I), and the soluble fraction (sup) and insoluble fraction (pre) of the IPTG-induced samples were loaded onto 10% SDS-PAGE. The arrow indicates the position of the fusion protein, and the molecular weight markers are labeled on the right side.

2.3.5 Detecting interactions between WW domain and a peptide with PPXY motif by FRET

RC (WW domain fused at the amino-terminus of ECFP) and PY (PPXY motif-containing peptide fused at the amino-terminus of EYFP) were expressed in E. coli, and the soluble fractions of total cell lysates were mixed together for FRET assay by using a fluorometer. RC was at a fixed concentration and titrated with different amounts of PY to form different mixtures. When PY concentration decreased, EYFP excitation peak for the mixtures also decreased (red lines in Figure 2.8). Similar patterns were also observed for PY samples with the same amounts of PY as the corresponding mixtures (black lines in Figure 2.8). By comparing the EYFP excitation peak of the mixtures to the EYFP excitation peak from the corresponding PY samples, no significant increase of EYFP excitation signals were observed for the mixtures (Figure 2.8). However, a decrease of ECFP signal was observed for the mixture with more PY (Mix1 and Mix2 in Figure 2.8). All the mixtures have the same amount of RC, and the excitation intensity of the mixture should be the same as the corresponding RC sample (Mix3 and RC in Figure 2.8). The decreases of ECFP signals for Mix1 and Mix2 in Figure 2.8 were possibly due to the interaction between WW domain and PPXY motif, resulting in energy transfer from ECFP to EYFP thereby quenching ECFP signal. However, the change of ECFP signal resulting from FRET is rather small.

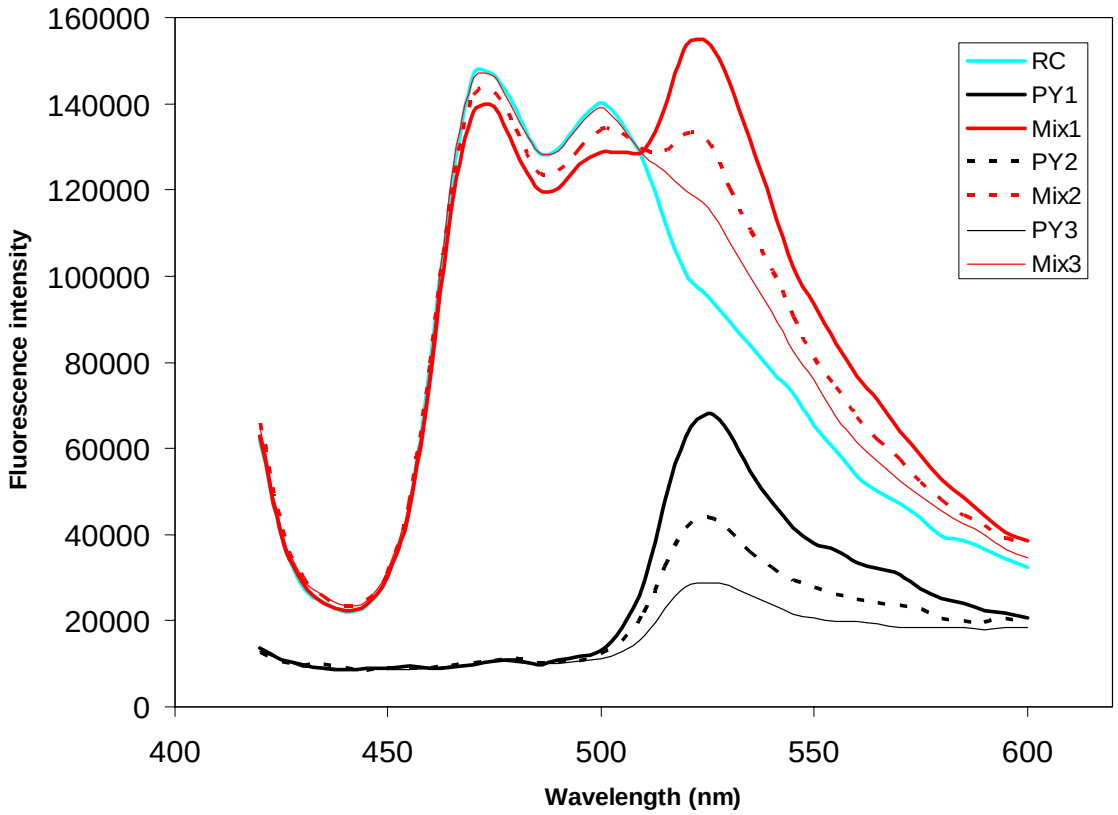


Figure 2.8 Detecting interactions between WW domain and a peptide with PPXY motif by mixing RC and PY.
The soluble fraction of expressed RC was used to mix with the soluble fraction of expressed PY. The concentration of RC was fixed and different amounts of PY were mixed with RC to obtain the mixtures. Mix1 contains more fusion protein PY than Mix2, and Mix2 contains more fusion protein PY than Mix3. The samples with only PY (PY1, PY2, and PY3) contain the fusion protein PY at the same concentrations as the corresponding mixture samples (Mix1, Mix2, and Mix3). RC contains fusion protein RC at the same concentrations as all the mixtures (Mix1, Mix2, and Mix3). All the emission spectra were obtained by fixing the excitation wavelength at 410 nm and scanning from 420 nm to 600 nm.

2.4 DISCUSSION

FRET signal was observed for EYFP-ECFP fusion proteins by using a fluorometer (Figure 2.5). Interactions between EYFP-tagged peptide with PPXY motif and ECFP-tagged WW domains were observed by fluorescence quenching of FRET donor ECFP (Figure 2.8). However, emission was observed for EYFP at the excitation wavelength used for FRET, which confounded FRET assay by using readout of sensitized emission of EYFP due to energy transfer from emission of ECFP. Correction for this effect has been documented, involving the acquisition of fluorescence images of the donor, the acceptor, and the sample for FRET assay (Gordon et al., 1998). However, this is not practical to scale up for high-throughput assays. FRET detection is also possible by monitoring the quenching of the donor emission (Bastiaens and Jovin, 1996). In this case, the emission from the donor without energy transfer is detected by first photobleaching the acceptor and then acquiring fluorescence emission signals from the donor. The difference of the donor emission fluorescence before and after photobleaching correlates with the number of interacting molecules between the donor and the acceptor. However, photobleaching usually requires 1-20 minutes and it will be more practical for fixed cells instead of live-cell measurement. FRET detection for live cells can be achieved by using fluorescence lifetime imaging (FLIM) that relies on the observation that FRET reduces the fluorescence lifetime of the donor molecule (Wouters and Bastiaens, 1999). Fluorescence lifetime can be determined by the fluorescence intensity change over time, which is time consuming. In conclusion, ECFP and EYFP can be used for FRET assay with the limitation of requiring either controls to monitor the direct emission from EYFP or time-consuming methods to determine FRET signals. The complications of FRET assay by ECFP and EYFP limit its utility in large-scale assays

and further improvements of the properties of ECFP and EYFP are necessary to increase the applicability of FRET in protein-protein interaction studies.

Chapter 3: Tandem Affinity Purification (TAP) of Transcription Factor Complexes

3.1 INTRODUCTION

Protein complex purification by using TAP tag was first introduced to identify a new component of U1 snRNP (Rigaut et al., 1999). TAP tag is comprised of two IgG-binding domains of protein A (ProA) from Staphylococcus aureus, a TEV protease recognition sequence, and a calmodulin-binding peptide (CBP). Accordingly, two-step purification can be employed to purify the TAP-tagged protein and its associated partners. The first step is based on the interaction between IgG and ProA tag to purify the protein complex by using IgG beads. Then TEV protease is used to cleave the protein complex off the beads. The second-step purification involves calmodulin beads that bind to CBP. The bound protein complex is eluted with EGTA (ethylene glycol tetra-acetic acid) and analyzed by SDS-PAGE and mass spectrometry. This method is used to purify protein complexes under mild conditions such that proteins in the complex are preserved in approximate stoichiometry. Two-step purification reduces non-specific bindings to the beads, therefore reducing contaminations in the purified protein complex. Additionally, by introducing TAP-tag into the genome in yeast cell to avoid protein overexpression typically occurring for plasmid-based protein expression, protein complexes can be purified under their physiological conditions. These advantages allow us to purify *bona fide* complexes, identify new interaction partners, and explore their functions in cells.

In yeast, there are over 200 transcription factors regulating gene expressions in cells (Lee and Young, 2000). Many transcription factors form a complex. For example, the general transcription factors form the holoenzyme with RNA polymerase II and mediator, responsible for transcription initiation (Myer and Young, 1998). The 2 MDa Swi/Snf complex consists of 11 proteins to regulate transcription activation by

remodeling chromatin structures (Peterson and Tamkun, 1995). By purifying transcription factor complexes to identify new components in the complexes, we could infer functional roles of the new components on the basis of their interaction partners whose functions are known.

I first genomically tagged 12 yeast transcription factors at their carboxyl-terminus with TAP tag: Rpd3p, Sin3p, Msn2p, Mcm1p, Pho2p, Hap2p, Hap3p, Phd1p, Mbp1p, Swi6p, Rpn4p, and Pdr3p. Later, TAP-tagged yeast library that covered about 75% of the proteome was available (Ghaemmaghami et al., 2003) and was also used in this study. In this chapter, I will present the complexes that were identified based on the TAP purifications of Rpd3p, Hap3p, and Swi6p, and attempt to decipher the functional roles for the new components in the purified complexes.

Rpd3p is a histone deacetylase that regulates silencing and transcription (Rundlett et al., 1996). By analyzing Mono-S profiles of deacetylase activity, Rpd3p has been shown to exist in a ~600 kDa complex (Rundlett et al., 1996). Later, purification of GST-Sin3p identified a 2 MDa complex, which also contained Rpd3p (Kasten et al., 1997). Several components were subsequently identified for this complex, such as Pho23p (Loewith et al., 2001), Sds3p (Alland et al., 2002).

Hap3p has been shown to form a complex with Hap2p and Hap5p (McNabb et al., 1995). In response to nonfermentable carbon sources, this complex binds to promoters with CCAAT boxes and activates the respiratory genes (Olesen et al., 1987; Pinkham and Guarente, 1985).

Swi6p interacts with Swi4p or Mbp1p to form two complexes: SBF (Swi4-Swi6 cell cycle box binding factor) or MBF (*MluI* binding factor), that regulate cell cycle G1-S transition (Koch et al., 1993). These two complexes are functionally redundant and cells with deletion of both *SWI4* and *MBP1* are inviable (Koch et al., 1993). However, SBP

binds genes predominantly functioning in budding as well as membrane and cell-wall biosynthesis, while genes bound by MBF are involved in DNA replication and repair (Iyer et al., 2001).

3.2 MATERIALS AND METHODS

3.2.1 Plasmid construction

pFA6a-TAP-KanMX6: TAP sequence was amplified from pBS1479 (Rigaut et al., 1999) by oligos GGCCTTAATTAACTCCATGGAAAAGAGAAGA and AAAAGGCGCGCCTCAGGTTGACTTCCCCGC. PCR fragment was gel-purified, digested with restriction enzymes PacI and AscI, and inserted into vector pFA6a-3HA-KanMX6 (Longtine et al., 1998). The final vector pFA6a-TAP-KanMX6 was confirmed by restriction enzyme digestion (PacI and AscI) and sequencing by using oligo TGTCGTTAGAACGCGGCTAC.

3.2.2 Strains

S288C (*MATα SUC2 mal mel gal2 CUP1 flo1 flo8-1 hap1*) was used as the parent strain to construct strains carrying genomically TAP-tagged transcription factors (Rpd3-TAP, Sin3-TAP, Msn2-TAP, Mcm1-TAP, Pho2-TAP, Hap2-TAP, Hap3-TAP, Phd1-TAP, Mbp1-TAP, Swi6-TAP, Rpn4-TAP, and Pdr3-TAP). The sequence containing TAP tag and *kan*^{r} selection marker was amplified from plasmid pFA6a-TAP-KanMX6 by using oligo pairs: one contains 45 or 50 nucleotides of the 3' end of the gene; another one contains 45 or 50 nucleotides that complement the 5' end of the 3'-untranslated sequence of the gene (Table 3.1). The PCR fragments were transformed into S288C and inserted between the last coding codon of the gene and the 5' end of the 3'-untranslated sequence by homologous recombination (Illustration 3.1) (Longtine et al., 1998). TAP-tagged

strains were confirmed by PCR and western blots by using rabbit PAP antibody (peroxidase anti-peroxidase complex) (Rockland Immunochemicals) against TAP tag.

TAP-tagged strains were also obtained from Open Biosystems, and they were constructed by inserting TAP-tag to the carboxyl-terminus of each gene in the parent strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) (Ghaemmaghami et al., 2003). Knockout strains were acquired from Research Genetics, and each ORF was deleted on the background strain BY4741 (Giaever et al., 2002).

Table 3.1 Oligo sequences for transcription factor TAP-tagging

Name	**Sequence (from 5' to 3')**
Rpd3F	CAATATGCGAGGGACCTACATGTTGAGCATGACAATGAATTCTATcggatccccgggttaattaa
Rpd3R	TTGTTTCACATTATTTATATTCGTATATACTTCCAACTCTTTTTTgaattcgagctcgtttaaac
Sin3F	ACGACTGGGAATACTGAATCTTCAGACAAGGGGGCTAAGATTCAAcggatccccgggttaattaa
Sin3R	CCCTGTCGTACTAAAGATTTTTGTTCTAAATCTAGTTAAAACTACgaattcgagctcgtttaaac
Msn2F	AATTTGTCGCAACACATCAAGACTCATAAAAAACATGGAGACATTcggatccccgggttaattaa
Msn2R	TTATCTTATGAAGAAAGATCTATCGAATTAAAAAAATGGGGTCTAgaattcgagctcgtttaaac
Mcm1F	GCTGCCTACCAACAATACTTTCAAGAACCGCAACAAGGCCAATACcggatccccgggttaattaa
Mcm1R	TCCTCTTAATGCTCGTCTATGAATTATATACGGAAATCGATAAGAgaattcgagctcgtttaaac
Pho2F	TTTTTGAAGAACACTAACGAACTAACTGACGAGCATAGATGGATAcggatccccgggttaattaa
Pho2R	AGTTGAAAATGCAATCGCAAAAAAAAAAAAAACAGAATTATTTTCAgaattcgagctcgtttaaac
Hap2F	CAGCCGCATGCTACTTCCACCGCAGCTGCAGCAGACAAAAAAACAcggatccccgggttaattaa
Hap2R	TTTTAGTTCTTTTAGGAATGATATTAACATTGGAATATTACAAAAgaattcgagctcgtttaaac
Hap3F	CTGAAGAATCAACTAATGTATGAGCAGGACGACGAAGAGGTGCCTcggatccccgggttaattaa
Hap3R	CTGACAGCTAGCAACTTTTGCGATCTACCACCTGGTTTTGTCTTCgaattcgagctcgtttaaac
Phd1F	ATCGATACTGAGGCCCAAACAAGTAGAGCAAAAAATGAATTATCAcggatccccgggttaattaa
Phd1R	GGTTCTTGTTCATAGAGCAAAGAGTTAACGGATTATGTTATGTGCgaattcgagctcgtttaaac
Mbp1F	GGCGCAGAACAGATCATCACAATCTCAAACGCGAATAGTCATGCAcggatccccgggttaattaa
Mbp1R	TTTCAGTATATGGATACATGTAAAGTTCCTCTATTTATGTATATTgaattcgagctcgtttaaac
Swi6F	ATTGACACTGACGAAATGCAAGATTTTTTAAAAAAGCATGCTTCAcggatccccgggttaattaa
Swi6R	CTTCAAATAAAGTCATAAAAGTTAATGCAATGAAATCACATGCCCgaattcgagctcgtttaaac
Rpn4F	gtcaaGAAGTTACAAAATTTGCAAAGGCTAATATTGGTTATGTCATGGGTcggatccccgggttaattaa
Rpn4R	ttttgTGTGAGGTTTTCTTCTTTTATCTCCTATATAATTTGTAACCTTAAgaattcgagctcgtttaaac
Pdr3F	CTGATTTATATCATACTCTGTGGAATGACAATACTTCATATCCCTTCTTAcggatccccgggttaattaa
Pdr3R	GTGTCCCATTTACTATGGTTATGCTCTGCTTCCCTATTTCTTTTGCGTTTgaattcgagctcgtttaaac

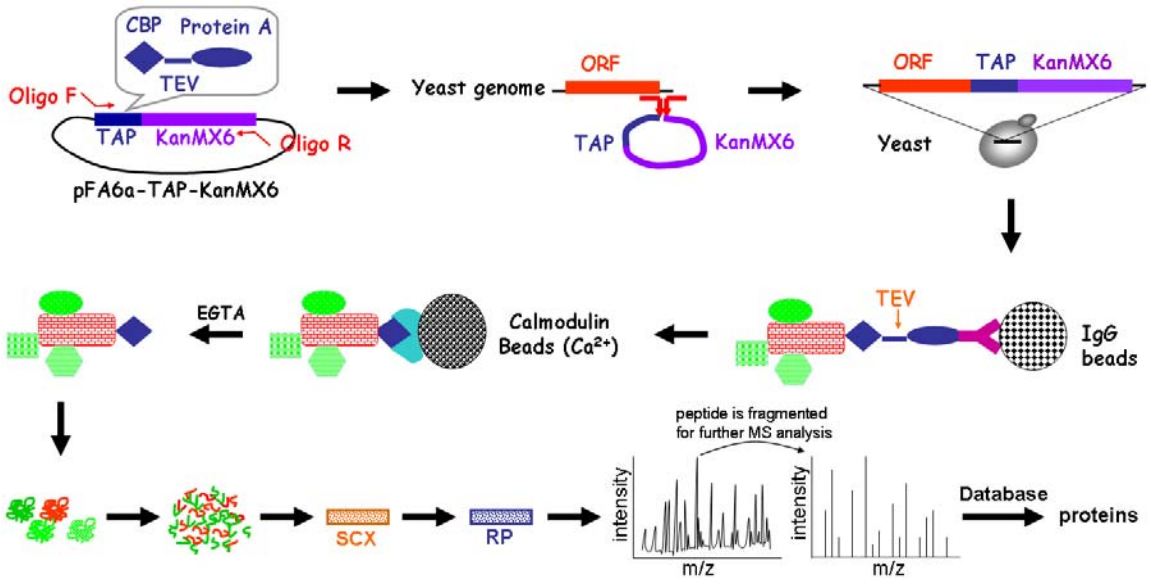


Illustration 3.1 TAP-tagged strain construction, TAP purification, and mass spectrometry. TAP tag that is composed of protein A domain, calmodulin binding motif, and TEV recognition sequence was amplified from plasmid (pFA6a-TAP-KanMX6) and inserted to carboxyl-terminus of a gene at its genomic locus in yeast to obtain a TAP-tagged strain (first row) that was used for affinity purification (second row). The purified sample was denatured and digested with trypsin to obtain the peptide mixture that was subjected to two-dimensional chromatography separation: strong cation column (SCX) and C18 reverse phase column (RP). The eluted peptides were analyzed with an online electrospray ionization ion trap mass spectrometer. The parent ion from the first mass spectrometry (MS) analysis was fragmented and further analyzed to obtain the peptide's sequence information that was utilized to obtain the protein identity by database searching (third row).

3.2.3 TAP purification

TAP purification was adapted from Séraphin Lab protocol (Illustration 3.1) (Rigaut et al., 1999). TAP-tagged strain was cultured to OD_{600} 1-2 in 4-10 L YPD (1% BactoYeast extract, 2% BactoPeptone, 2% dextrose) at 30°C. The following steps were performed at 4°C unless otherwise indicated. Cells were spin down by centrifugation, washed once with cold water, and lysed with beadbeater (Biospec) in cold lysis buffer (6 mM Na_2HPO_4, 4 mM NaH_2PO_4, 0.5% NP-40, 150 mM NaCl, 2 mM EDTA, 50 mM NaF, Na_3VO_4) with protease inhibitors (2 ug/ml leupeptin, 2 ug/ml aprotinin, 1 ug/ml bestatin, 1 ug/ml pepstatin A, 1 mM PMSF). Cell lysates were clarified by centrifugation at 20,000g. The supernatant was incubated with IgG Sepharose beads (Amersham) for 2 hours with rotating. The beads were washed with 30 ml buffer (10 mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40) and then 10 ml TEV cleavage buffer (10 mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40, 0.5 mM EDTA, 1 mM DTT). The beads were incubated with 1 ml TEV cleavage buffer and 500 U TEV (Invitrogen) for 2 hours at 16°C with rotating. The eluate was collected. The beads were washed with 1 ml TEV cleavage buffer and the eluate was combined with first eluate. 6 ml CBB (10 mM Tris pH 8.0, 150 mM NaCl, 1 mM MgAcetate, 1 mM Imidazole, 2mM $CaCl_2$, 10 mM β-mercaptoethanol, 0.1% NP-40) and 6 ul $CaCl_2$ were added into the eluate, and incubated with Calmodulin beads (Amersham) for 1 hour. The beads were washed twice with 10 ml CBB and once with 10 ml CBB with 0.02% NP-40. The bound proteins were eluted with 1 ml CEB (10 mM Tris pH 8.0, 150 mM NaCl, 1 mM MgAcetate, 1 mM Imidazole, 0.02% NP-40, 20 mM EGTA, 10 mM β-mercaptoethanol). The eluate was split into two halves. Proteins in the eluate were precipitated with 25% cold trichloroacetic acid. The pellets were washed with 100% cold acetone. One pellet was resuspended in 100 mM Tris buffer (pH 8.0), and

digested with proteomic grade trypsin (Sigma) for 24 hrs. Another pellet was resuspended in Laemmli buffer for SDS-PAGE.

3.2.4 Mass spectrometry

Tryptic peptide mixtures were first loaded onto a strong cation exchange column (SCX) and eluted with 11 salt concentrations: 0, 20, 40, 60, 80, 100, 150, 200, 300, 500, and 900 mM ammonium chloride. Each salt eluate was loaded onto a reverse phase C18 column and peptides were eluted with a 60-minute gradient from 5% to 45% acetonitrile (ACN) and analyzed online by an electrospray ionization (ESI) ion trap mass spectrometer (ThermoFinnigan DecaXP Plus). For each parent ion spectrum, 3 tallest peaks were selected and fragmented by collision-induced dissociation (CID) with helium gas to produce MS/MS spectra. Data were searched against a protein sequence database of *Saccharomyces cerevisiae* by SEQUEST (Bioworks, Thermo) (Illustration 3.1). The peptide cross correlation score (Xcor) 2.5 was used to filter SEQUEST results to obtain the positive identifications. Each identified protein was ordered by rank based on its protein score. Higher protein score gives higher rank representing relatively higher abundance in the sample. Accordingly, proteins identified from reference samples (total cell lysate or TAP purification of the wild type strain) were rank ordered based on the protein scores. By comparing the rank of each protein in the purified sample to the reference samples, the proteins enriched in the purified sample were considered as true positive identifications.

3.2.5 Total RNA extraction

Cells were cultured to OD_{600} ~1.0 under the indicated conditions, centrifuged for 5 minutes at 3,000g, frozen in liquid nitrogen, and stored in -80°C. Total RNA was extracted by hot acidic phenol method (DeRisi et al., 1997). Cell pellet was resuspended

in AE buffer (50 mM sodium acetate pH 5.2, 10 mM EDTA) and equal volume of acidic phenol (Invitrogen). SDS was added to a final concentration of 0.8%. The mixture was incubated at 65°C for 1 hour with agitation every 10 minutes, and immediately incubated on ice for 10 minutes. After centrifugation at 3,000g for 10 minutes, the aqueous phase was extracted with equal volume of acidic phenol, followed by chloroform extraction. Heavy Phase Lock Gel (Eppendorf) was utilized to separate the organic and aqueous phase. RNA was precipitated with two volume of 100% ethanol and one tenth volume of sodium acetate (3 M, pH 5.2). RNA pellet was washed once with cold 70% ethanol, dried, resuspended in diethylpyrocarbonate (DEPC) water, and stored in -80°C.

3.2.6 DNA microarray

15 ug of total RNA was utilized for reverse transcription by SuperScript II reverse transcriptase (Invitrogen), oligo dT (5'-TTTTTTTTTTTTTTTTTTTTVN-3'), and dNTP mix containing amino-allyl dUTP. cDNA was purified by by MinElute column (Qiagen) and labeled with Cy5 or Cy3 mono-NHS-ester dyes (Amersham) in the dark at room temperature (RT). Unincorporated dyes were removed by MinElute column. Cy3-labeled reference cDNAs and Cy5-labeled sample cDNAs were mixed together with yeast tRNA and polyA RNA, hybridized onto yeast whole genome microarray constructed by Vishy Iyer lab at 65°C overnight. Arrays were washed, centrifuged to dry, coated with Dye-Saver2 (Genisphere), and scanned with GenePix 4000A or B scanner (Axon Instruments) at Cy3 (532 nm) and Cy5 (635 nm) channels. Scanned images were fitted onto grids and analyzed with GenePix 4.0 or 5.0 (Axon Instruments). The resulting data were uploaded into Longhorn Array Database (Killion et al., 2003) for data processing. The medium of intensity for each gene was used to calculate the base 2 logarithm ratio of the sample to the reference. Hierarchical clustering was used to cluster the genes (Eisen et al., 1998).

3.3 RESULTS

3.3.1 Identification of new components of histone deacetylase complex B (HDB)

Strain carrying Rpd3-TAP allele was utilized to purify HDB from 4 L culture by TAP purification. Half of the purified protein sample was separated by 10% SDS-PAGE and stained with Coomassie Blue followed by silver staining. Very faint bands were observed by Coomassie Blue staining (data not shown), whereas clear bands were observed by silver staining (Figure 3.1A). Half of the sample was analyzed by mass spectrometry and MS rankings for identified proteins from Rpd3-TAP purified sample were compared to MS rankings for proteins from the total cell lysate. The known components of HDB were identified including Rpd3p, Sin3p, Sds3p, and Pho23p (Figure 3.1B). Several proteins not known in this complex were identified, such as Fun19p, Dep1p, Ybr095cp, Ume1p, Ymr075wp, Hap2p, and Ymr263wp. Their exact roles in HDB complex remain to be determined.

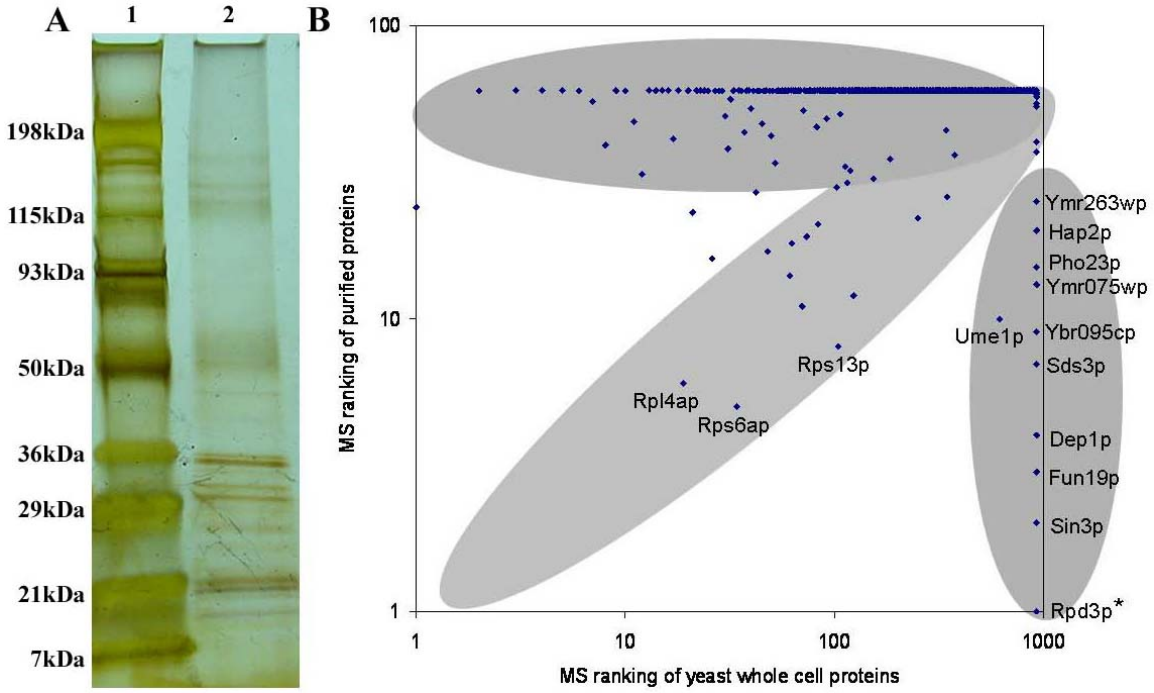


Figure 3.1 Affinity purification of HDB complex by using Rpd3p as bait protein.
(A) Half of the purified sample was loaded onto a 10% SDS-PAGE gel and stained with silver (lane 2). The molecular weight markers were in lane 1 and labeled on the left side.
(B) Half of the purified sample was analyzed by mass spectrometry. The mass spectrometry rankings for proteins in the purified sample (y axis) were plotted against the mass spectrometry rankings for proteins identified from the whole cell lysate (x axis). The bait protein (labeled with star) for the purification, Rpd3p, was ranked number 1 in the purified sample and ranked 900 in the whole cell lysate sample, showing extremely strong enrichment in the purified sample. Three areas were highlighted with gray color: proteins with strong enrichment in the purified sample on the right side representing the specific purified components in the complex; proteins enriched in the whole cell lysate on the top; proteins in both purified sample and the whole cell lysate on the diagonal representing most contaminants during purification.

3.3.2 Identification of Gln3p as a new component in Hap2/3/5 complex

In an effort to purify Hap2/3/5 complex by using Hap3-TAP as the bait, Gln3p was identified as a new component (Figure 3.2A). Additionally, Gln3p was also identified in the TAP purifications by using either Hap2-TAP or Hap5-TAP as the bait proteins (Figure 3.2B-C). However, for TAP purification by using Gln3-TAP as the bait protein, the known Gln3p interacting protein Tor1p was identified (Carvalho and Zheng, 2003), and additional identified proteins included Kog1p and Tco89p, but not the Hap2/3/5 complex (Figure 3.2D). Tor1p is a protein kinase that regulates cell growth in response to nutrients and stresses (Lorberg and Hall, 2004). Kog1p has been shown to interact with Tor1p to form Tor complex 1 (TORC1) (Loewith et al., 2002). Tco89p was an uncharacterized ORF by then and later was independently identified by another group as a component of TORC1 (Reinke et al., 2004). Here I showed that Gln3p is also part of TORC1 possibly by interacting with Tor1 (Carvalho and Zheng, 2003). Therefore, Gln3p might interact with different complexes in the cell (Illustration 3.2), which can be partially supported by protein localization information: Gln3-GFP localizes in both nucleus and cytoplasm, whereas Hap2/3/5 complex localizes in the nucleus and the components of TORC1 localize in the cytoplasm (Huh et al., 2003). One possible reason that Hap2/3/5 complex was not identified by TAP purification of Gln3-TAP could be the low abundance of Hap2/3/5-Gln3 complex in the nucleus of cell. Another possibility might be the TAP tag on the carboxyl-terminus of Gln3p that could interfere with the interaction between Hap2/3/5 and Gln3p.

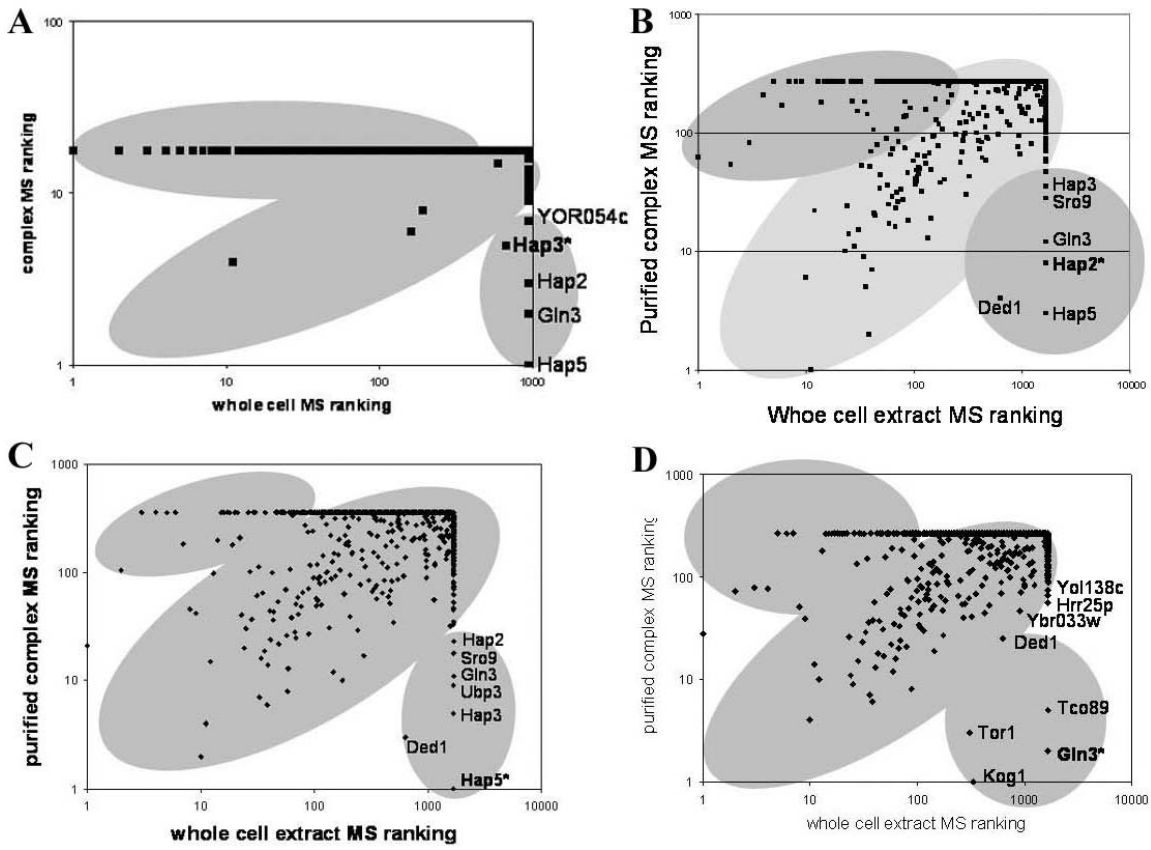


Figure 3.2 TAP purifications of Hap2/3/5 complex and Gln3p interacting proteins.
Each plot demonstrates the enrichment of proteins in the purified sample (MS rankings on the y axis) relative to the whole cell lysate sample (MS rankings on the x axis). Highly enriched proteins in the purified sample locate in the bottom-right area highlighted with gray color, representing specific proteins in the purified complex. The proteins locating at the diagonal are mostly contaminants in the purification. The bait proteins are labeled with stars, and they are top ranked in the purified samples. Gln3p was observed in the purifications by using Hap3-TAP (A), Hap2-TAP (B), and Hap5-TAP (C) as bait proteins. However, Hap2/3/5 complex was not observed in Gln3-TAP purification (D). Instead, TORC1 components were identified in Gln3-TAP purification sample (D).

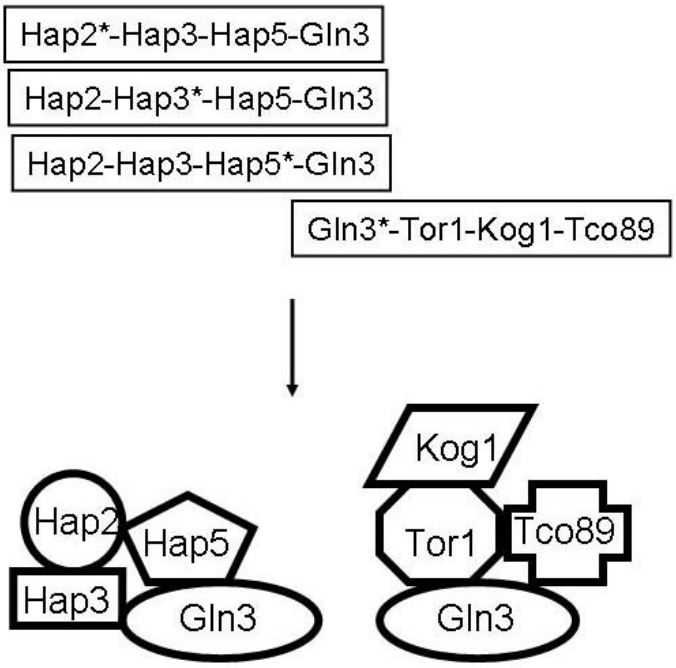


Illustration 3.2 Gln3p interacts with Hap2/3/5 complex and TORC1.
Based on TAP purifications (Figure 3.2) by using Hap2-TAP, Hap3-TAP, Hap5-TAP, and Gln3-TAP as the bait proteins (labeled with star), a model is proposed for two complexes: Hap2/3/5-Gln3 complex and TORC1-Gln3 complex. This protein complex model reflects neither the binary interaction information among the components, nor the stoichiometry information for the components in the complex.

3.3.2.1 Interaction between Hap2/3/5 complex and Gln3p is DNA-independent

Hap2/3/5 complex binds to CCAAT box-containing sequences (McNabb et al., 1995), and Gln3p binds to sequences that contain GATAAGA motif (Harbison et al., 2004). If the promoter sequence of a gene contains both sequence motifs, the observed association of Hap2/3/5 and Gln3p in affinity purification is likely due to the promoter sequence that brings Hap2/3/5 and Gln3p together, instead of due to direct protein-protein interactions. In yeast cells, the promoter sequence of *GDH1* gene has been shown to contain motifs for multiple transcription factors, including Gln3p, Hap2/3/5 complex, Gcn4p, and Leu3p (Riego et al., 2002). In a genome-wide location analysis, there are about 20 genes whose promoter sequences associate with both Hap2p and Gln3p (Harbison et al., 2004). Therefore, in order to distinguish DNA-dependent association and DNA-independent association, ethidium bromide (EtBr) was utilized to disrupt protein-DNA interactions (Lai and Herr, 1992) during affinity purification by using Hap5-TAP as the bait. Gln3p was identified in affinity purifications either with or without EtBr (Figure 3.3), suggesting that association of Hap2/3/5 with Gln3p was DNA-independent.

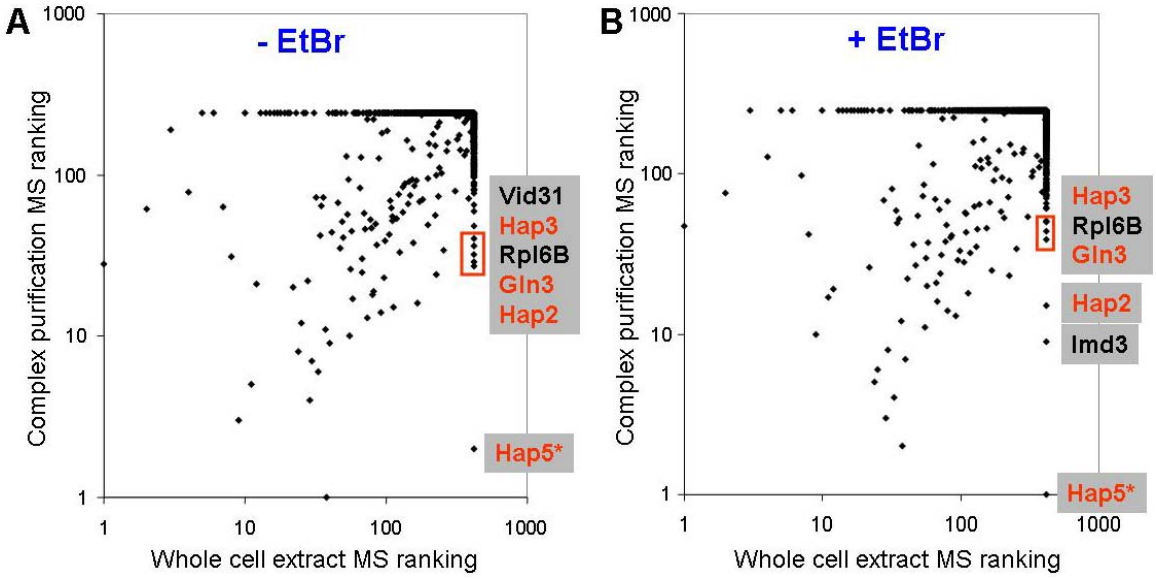


Figure 3.3 Affinity purification of Hap2/3/5 complex with or without EtBr.
Hap5-TAP was used as the bait (labeled with star) for affinity purification. The known components (Hap2p and Hap3p), and the new component (Gln3p) were identified from both purifications without (A) or with (B) EtBr.

3.3.2.2 Gln3p co-immunoprecipitated with Hap2/3/5 complex

To further confirm that Gln3p interacts with Hap2/3/5 complex, Gln3p, Hap2p, Hap3p, Hap5-TAP were co-expressed in yeast cells, immunoprecipitation was employed by using IgG beads to pull down Hap5-TAP. Gln3p was detected by Western blot (Figure 3.4A), suggesting that Gln3p associated with Hap5-TAP. Mass spectrometry analysis also showed that Gln3p was in the pull-down sample (Figure 3.4B). These data indicated that Gln3p physically associates with Hap2/3/5 complex.

3.3.2.3 Deletion of both HAP3 and GLN3 did not show synthetic phenotype

In order to decipher the functional relationship between Hap2/3/5 complex and Gln3p, the double deletion strain *hap3gln3Δ* was constructed. *hap3Δ*, *gln3Δ*, and *hap3gln3Δ* mutants exhibited slightly slow growth phenotypes compared to wild type cells, and the growth rate of the double deletion mutant showed little difference from the single deletion mutants in liquid YPD medium at 30°C (Figure 3.5A). Additionally, there were no obvious morphology defects for either the single deletion mutants or the double deletion mutant in YPD medium at 30°C (Figure 3.5B). Hap2/3/5 and Gln3p are transcription factors that activate genes in response to carbon and nitrogen sources, respectively. Growth assays were performed for the single deletion mutants and the double deletion mutant under different carbon and nitrogen sources. On YPD plate, *hap3Δ* and *hap3gln3Δ* grew slower than *gln3Δ* and wild type, and exhibited star-shape white colonies (Figure 3.6) that are characteristics of mitochondrial incompetence (Devin and Koltovaya, 1987), whereas *gln3Δ* and wild type cells exhibited round-shape gray colonies (Figure 3.6). The different growth phenotypes observed in YPD liquid cultures and on YPD plates for *hap3Δ*, *gln3Δ*, and *hap3gln3Δ* mutants are likely due to the growth conditions: cells obtain better nutrients in shaked liquid cultures than on plates, and the

growth rates in liquid cultures were measured in logarithm phase. *hap3Δ* and *hap3gln3Δ* did not grow in rich medium with glycerol as the carbon source (Figure 3.6), consistent with the role of Hap2/3/5 complex in respiratory regulation (Pinkham and Guarente, 1985). The mutants and wild type strain grew similarly in media with different nitrogen sources (Figure 3.6).

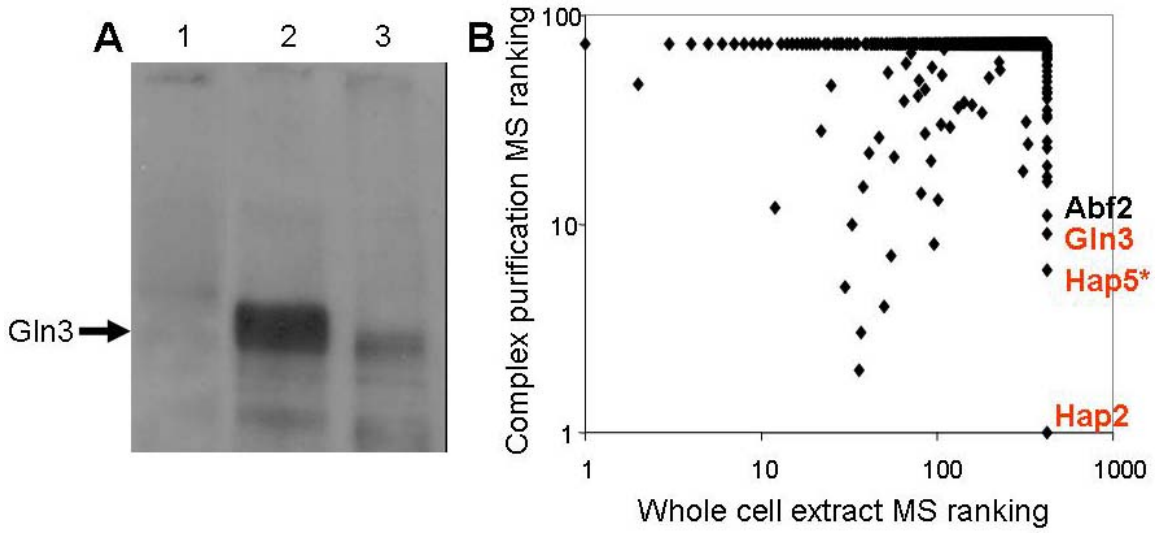


Figure 3.4 Gln3p co-immunoprecipitated with Hap5-TAP.
Gln3p, Hap2p, Hap3p, and Hap5-TAP were co-expressed under GAL1/10 promoter controls in yeast cells. 300 ml cultures were used for IgG immunoprecipitation. Half of the pull-down sample was used for detecting Gln3p by using the antibody against Gln3p (lane 3 in panel A), and the other half of the sample was used for mass spectrometry analysis (B). Gln3p was not detected in the negative control by IgG immunoprecipitation of samples obtained from cells co-expressing Gln3p, Hap2p, Hap3p, and Hap5p (lane 1 in panel A). Gln3p was detected in the total lysate from cells co-expressing Gln3p, Hap2p, Hap3p, and Hap5p-TAP (lane 2 in panel A). Gln3p was also detected in the co-immunoprecipitation sample by using IgG beads to pull down Hap5-TAP from lysate containing co-expressed Gln3p, Hap2p, Hap3p, and Hap5p-TAP (lane 3 in panel A). Mass spectrometry analysis also showed that Gln3p was in the co-immunoprecipitation sample.

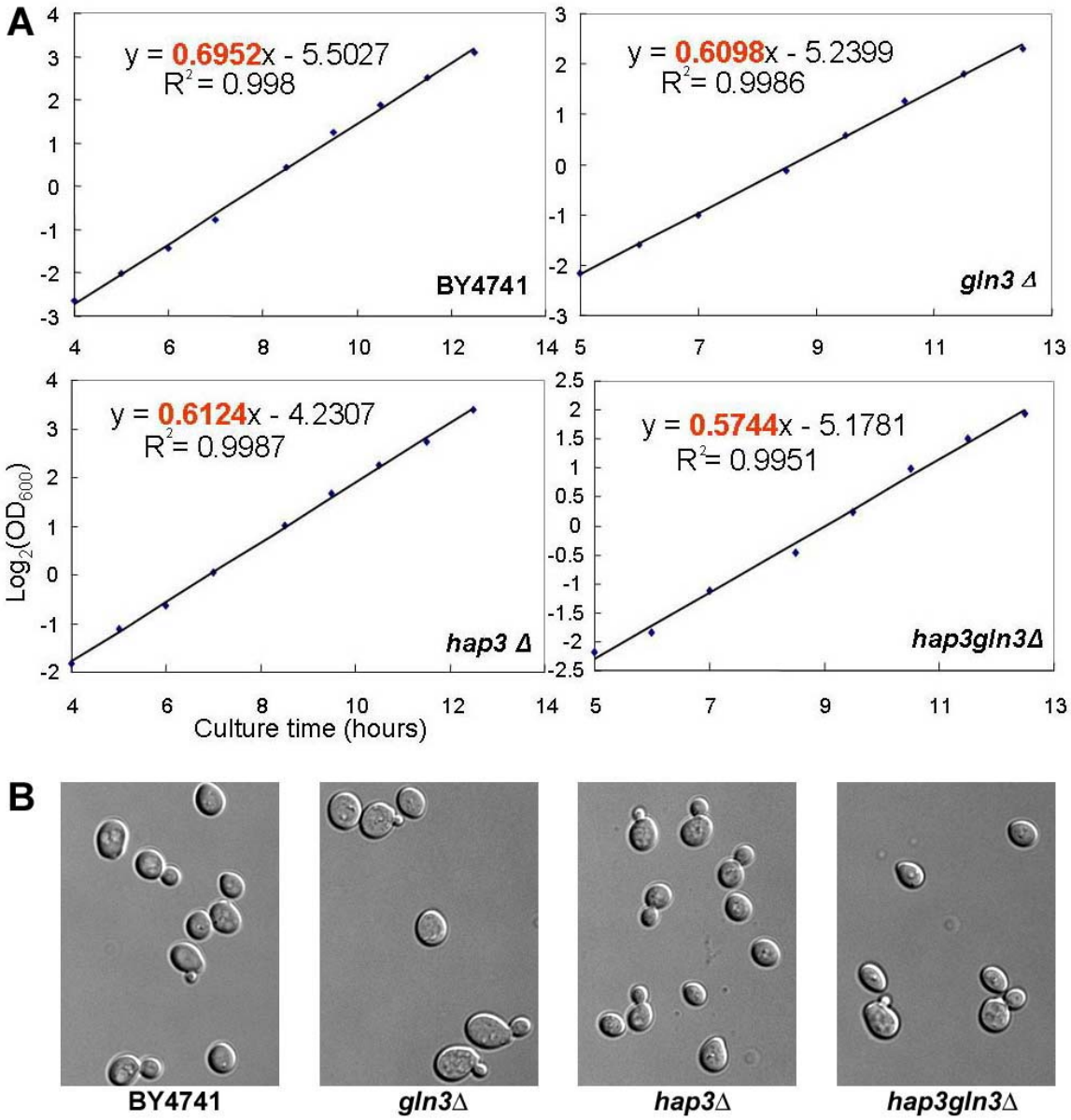


Figure 3.5 Growth rates and morphology phenotypes of hap3Δ, gln3Δ, hap3gln3Δ.
(A) Wild type cells and mutants were cultured in YPD at 30°C. OD_{600} was measured at nine different time points, and the base 2 logarithm of OD_{600} was calculated to plot the growth curve as a function of culture time. The slope of each linear regression fit line represents the growth rate (red color number).
(B) Cells were observed under the microscope.

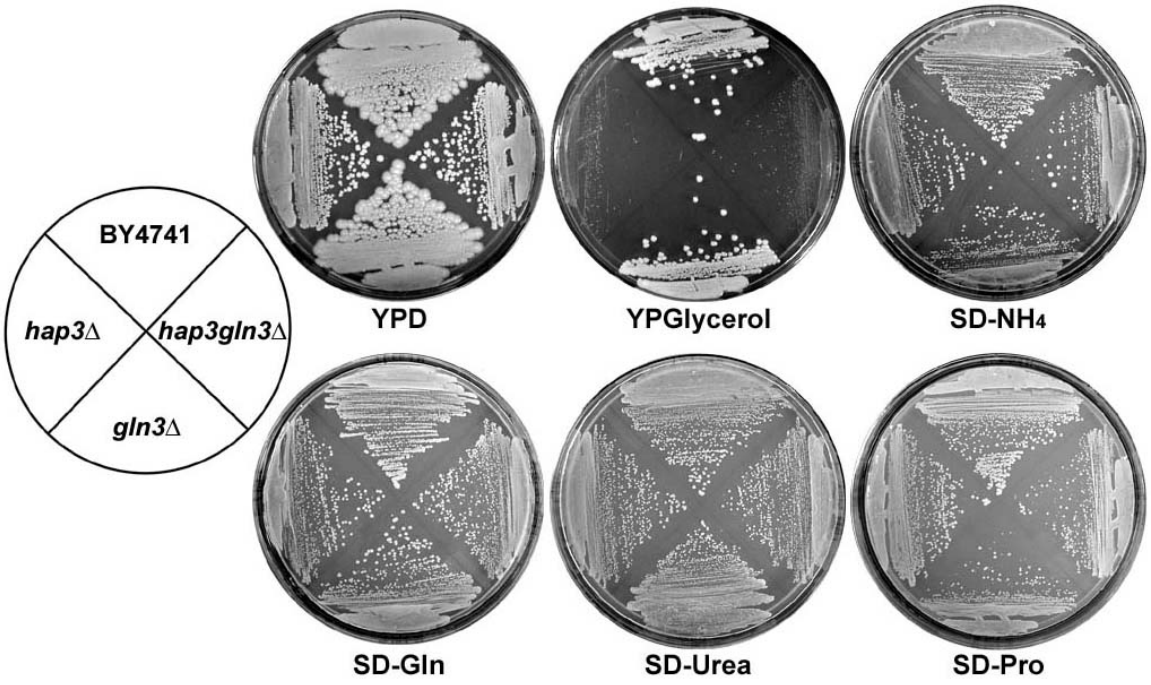


Figure 3.6 Growth assays for BY4741, hap3Δ, gln3Δ, and hap3gln3Δ under different nutrient conditions.
Cells were streaked onto plates with different carbon and nitrogen sources: rich medium with 2% dextrose as carbon source (YPD); rich medium with 5% glycerol as carbon source (YPGlycerol); synthetic complete medium with dextrose as carbon source and ammonium as nitrogen source (SD-NH4); synthetic complete medium with dextrose as carbon source and glutamine as nitrogen source (SD-Gln); synthetic complete medium with dextrose as carbon source and urea as nitrogen source (SD-Urea); synthetic complete medium with dextrose as carbon source and proline as nitrogen source (SD-Pro). All the plates were cultured at 30°C.

3.3.2.4 Identification of genes regulated by Hap2/3/5 complex and Gln3p

Hap2/3/5 complex and Gln3p are transcription factors that regulate gene expressions in response to nutrients. In order to elucidate the functional relationship between these two transcription factors, gene expression profiling by DNA microarray was employed for strains with deletion of *HAP3*, *GLN3*, or both genes in rich medium (YPD). 49 genes were identified to be up-regulated or down-regulated at least two-fold in at least one mutant (Figure 3.7). Functional classification analysis by FunSpec (Functional Specification) (Robinson et al., 2002) based on MIPS (Munich Information Center for Protein Sequences) functional categories showed enrichment of DNA metabolism, ion transporter, and TCA genes for the up-regulated genes, and nitrogen, sulfur, and phosphate metabolism genes for the down-regulated genes (Table 3.2 and 3.3). Most of those genes were regulated by either *HAP3* or *GLN3* exclusively, suggesting their different functional roles in the cell. Several genes were regulated by both genes such as *GDH1*, *FET3*, *ACO1*, *AHP1*, and *HSP30* (Figure 3.7). *GDH1* is a known gene regulated by both Hap2/3/5 complex and Gln3p (Riego et al., 2002). *HSP30* was up-regulated in *hap3Δ* mutant and down-regulated in *gln3Δ* mutant. However, in *hap3gln3Δ* mutant, *HSP30* expression did not show difference from wild type cells (Figure 3.7), which was possibly because the increased level of *HSP30* mRNA in *hap3Δ* compensated the decreased level of *HSP30* mRNA. The limited number of genes co-regulated by both transcription factors suggested that functional interaction between these two transcription factors was rather weak, at least in rich medium at 30°C. The data also showed that Hap2/3/5 complex and Gln3p function constitutively in carbon and nitrogen rich medium.

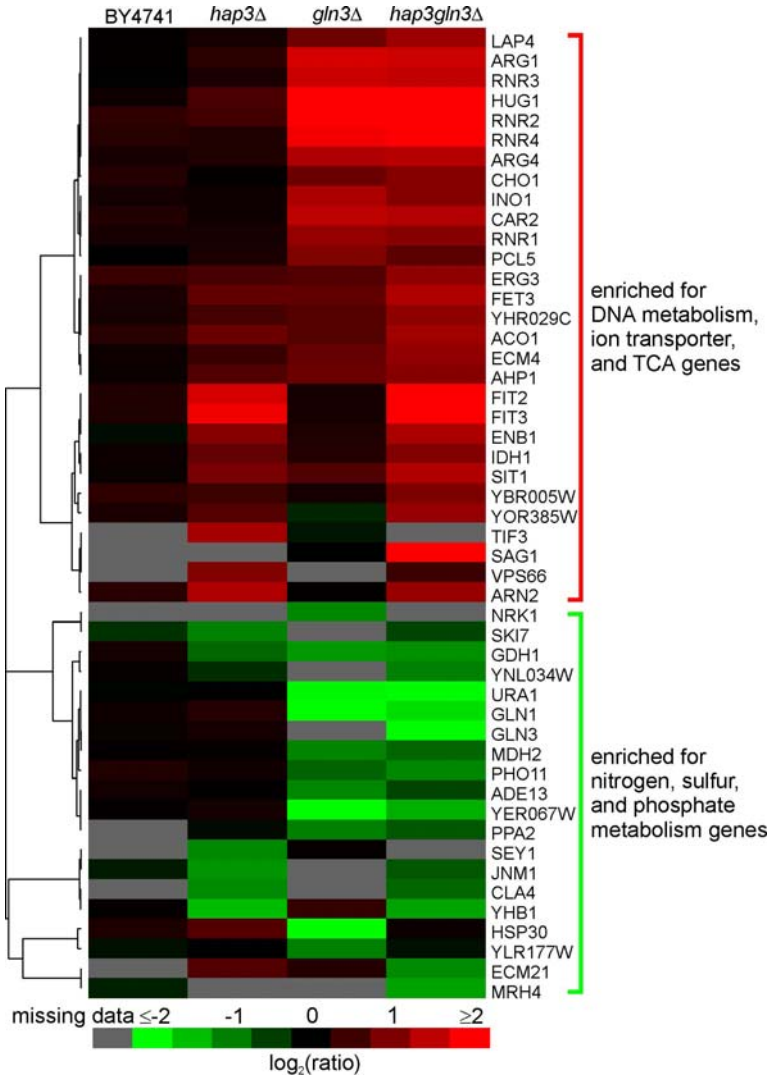


Figure 3.7 Gene expression profiling by DNA microarray for hap3Δ, gln3Δ, and hap3gln3Δ in rich medium.
Total RNAs from BY4741, *hap3Δ*, *gln3Δ*, and *hap3gln3Δ* were reverse transcribed to cDNAs labeled with Cy5 (red signal) that were hybridized with BY4741 control cDNA sample that was labeled with Cy3 (green signal). The base 2 logarithm ratio of the medium intensity of red signal to that of green signal was used for hierarchical clustering analysis.

Table 3.2 MIPS functional classification for genes with 2-fold up-regulation in at least one mutant of *hap3Δ*, *gln3Δ*, and *hap3gln3Δ*.

Primary	Secondary	p-value	k	f
	deoxyribonucleotide metabolism	1.19E-07	4	11
	drug transporters	0.000527939	3	35
	DNA synthesis and replication	0.000845882	4	94
METABOLISM		0.00167783	12	1066
	cytoplasm	0.00285609	8	554
	nucleotide metabolism	0.00447964	4	148
	heavy metal ion transporters (Cu, Fe, etc.)	0.00492549	2	23
	DNA processing	0.00534154	5	251
	tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)	0.00580716	2	25
	DNA repair	0.00750973	3	88
CELL RESCUE, DEFENSE AND VIRULENCE		0.00819209	5	278

p-value: the probability that the intersection of given list with any given functional category occurs by chance
k: number of genes with 2-fold up-regulation in the given category
f: number of genes total in the given category

Table 3.3 MIPS functional classification for genes with 2-fold down-regulation in at least one mutant of *hap3Δ*, *gln3Δ*, and *hap3gln3Δ*.

Primary	Secondary	p-value	k	f
	phosphate utilization	0.000860577	2	14
	nitrogen and sulfur metabolism	0.0011691	3	67
METABOLISM		0.00323108	9	1066
	phosphate metabolism	0.00481507	2	33
	nitrogen and sulfur utilization	0.00635006	2	38

p-value: the probability that the intersection of given list with any given functional category occurs by chance
k: number of genes with 2-fold down-regulation in the given category
f: number of genes total in the given category

It has been shown that Hap2/3/5 complex and Gln3p regulate gene expressions under certain conditions. For example, Gln3p will move from cytoplasm to nucleus to activate genes required for nitrogen utilization upon nitrogen limitation or in poor nitrogen sources such as urea or proline (Courchesne and Magasanik, 1988). In order to elucidate the functional roles of Hap2/3/5 complex and Gln3p upon nutrient limitation, *hap2Δ*, *hap3Δ*, *hap4Δ*, *hap5Δ*, *gln3Δ*, *hap3gln3Δ*, and wild type cells were treated with rapamycin to induce a nutrient starvation state (Heitman et al., 1991). By comparing gene expression levels of mutants to that of wild type cells under nutrient starvation condition, expression levels of 121 genes changed at least 2-fold in at least one of the mutants under rapamycin condition (Figure 3.8). The expression levels of most of these genes changed dramatically in *gln3Δ* cells, but not in the *hap* deletion mutants (Figure 3.8), consistent with the regulatory role of Tor1p on Gln3p (Beck and Hall, 1999). The up-regulated genes exhibited enrichment in carbon and nucleotide metabolism, and stress response pathway, whereas the down-regulated genes showed enrichment in amino acid, nitrogen, sulfur, and nucleotide metabolism and transport (Table 3.4 and 3.5). A small group of genes were co-regulated by both Hap complex and Gln3p (Figure 3.8). For example, genes involved in arginine biosynthesis pathway were down-regulated in most mutants: *ARG1*, *ARG4*, and *ARG5,6* (Figure 3.8). *BTN2* that modulates arginine uptake was up-regulated in all mutants (Figure 3.8). *HSP30* was also up-regulated in all mutants upon rapamycin treatment, in contrast with the observation that it was down-regulated in *gln3Δ* without rapamycin treatment.

Hap2/3/5 complex binds to CCAAT box on promoter region and Hap4p provides the activation function for the complex (Forsburg and Guarente, 1989). However, several genes were down-regulated in *hap2Δ*, *hap3Δ*, and *hap5Δ*, but not in *hap4Δ*: *GDH1*,

GDH3, *YHB1*, and *YIL165C* (Figure 3.8), suggesting that Hap2/3/5 complex could activate some genes and function in a Hap4p-independent manner.

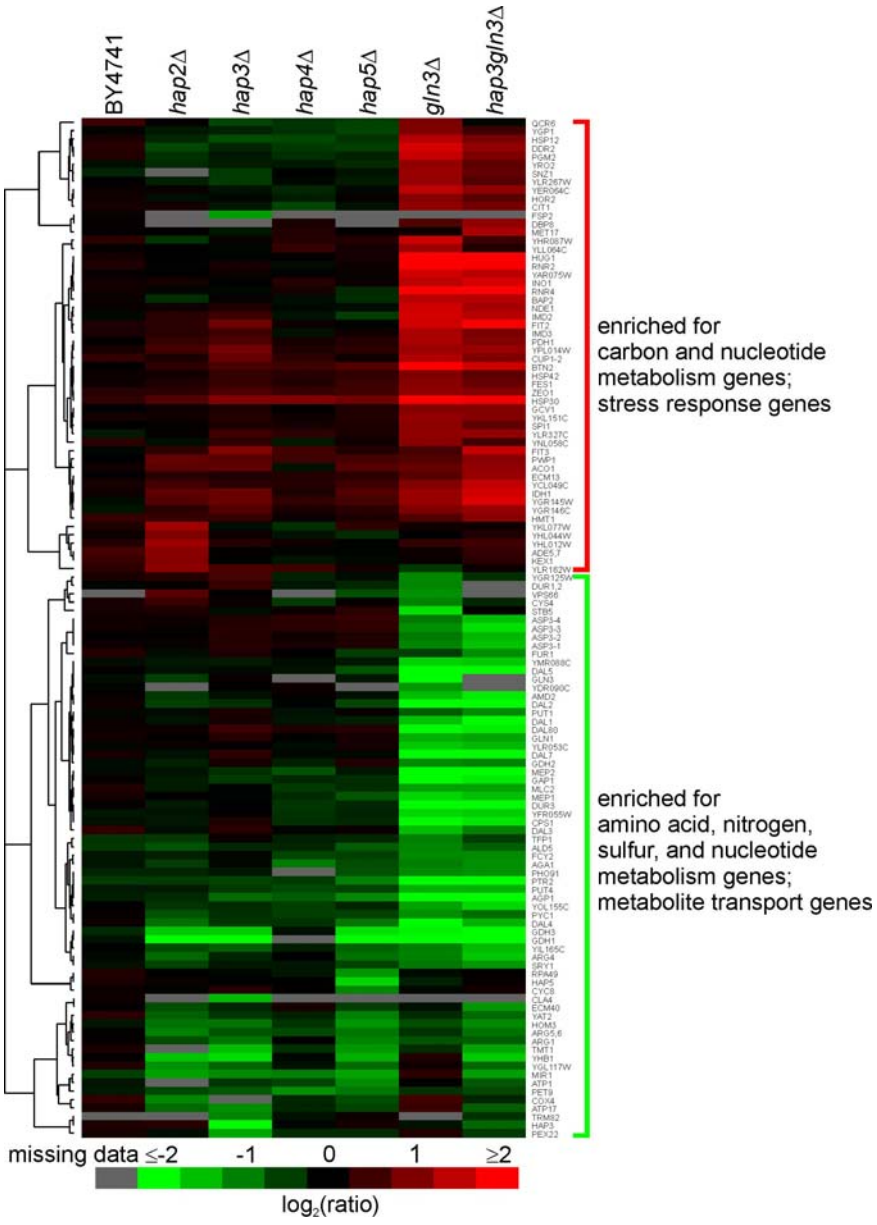


Figure 3.8 Gene expression profiling by DNA microarray for hap2Δ, hap3Δ, hap4Δ, hap5Δ, gln3Δ, and hap3gln3Δ upon rapamycin treatment.
Total RNAs from BY4741 and the mutants were reverse transcribed to cDNAs labeled with Cy5 (red signal) that were hybridized with BY4741 control cDNA sample that was labeled with Cy3 (green signal). The base 2 logarithm ratio of the medium intensity of red signal to that of green signal was used for hierarchical clustering analysis.

Table 3.4 MIPS functional classification for genes with 2-fold up-regulation in at least one mutant of *hap2Δ*, *hap3Δ*, *hap4Δ*, *hap5Δ*, *gln3Δ*, and *hap3gln3Δ*.

Primary	**Secondary**	**p-value**	**k**	**f**
	stress response	0.000109386	8	175
	C-compound and carbohydrate utilization	0.000336407	9	261
CELL RESCUE, DEFENSE AND VIRULENCE		0.000535139	9	278
	purine ribonucleotide metabolism	0.000564715	4	45
	tricarboxylic-acid pathway (citrate cycle, Krebs cycle, TCA cycle)	0.00121642	3	25
	nucleotide metabolism	0.0015902	6	148
METABOLISM		0.00255725	18	1066
	deoxyribonucleotide metabolism	0.00381368	2	11
ENERGY		0.00549969	7	252
	C-compound and carbohydrate metabolism	0.00829697	9	415

p-value: the probability that the intersection of given list with any given functional category occurs by chance
k: number of genes with 2-fold down-regulation in the given category
f: number of genes total in the given category

Table 3.5 MIPS functional classification for genes with 2-fold down-regulation in at least one mutant of *hap2Δ*, *hap3Δ*, *hap4Δ*, *hap5Δ*, *gln3Δ*, and *hap3gln3Δ*.

Primary	**Secondary**	**p-value**	**k**	**f**
	amino acid metabolism	1.00E-14	21	204
	nitrogen and sulfur utilization	7.76E-14	11	38
	nitrogen and sulfur metabolism	1.05E-13	13	67
	amino acid degradation (catabolism)	1.44E-12	10	35
METABOLISM		6.37E-12	36	1066
TRANSPORT FACILITATION		1.60E-08	17	312
	plasma membrane	2.94E-07	11	145
	anion transporters (Cl, SO4, PO4, etc.)	2.14E-06	5	21
	amino acid biosynthesis	3.74E-06	9	118
SUBCELLULAR LOCALISATION		2.17E-05	41	2256
	cellular import	8.47E-05	7	100
	purine ribonucleotide metabolism	0.000105393	5	45
	nucleotide metabolism	0.000160599	8	148
	ion transporters	0.000167109	6	78
	homeostasis of anions	0.000309441	3	13
CELLULAR TRANSPORT AND TRANSPORT MECHANISMS		0.000589185	14	494
	cell wall	0.000670115	4	38
	other cation transporters (Na, K, Ca , NH4, etc.)	0.000670115	4	38
	amino acid transport	0.00177507	3	23
	amino-acid transporters	0.00227038	3	25
	homeostasis of phosphate	0.00228423	2	7
	allantoin and allantoate transporters	0.00386206	2	9
	cation transporters	0.0039415	4	61
	phosphate transport	0.00479436	2	10
	nucleotide transport	0.00943168	2	14

p-value: the probability that the intersection of given list with any given functional category occurs by chance
k: number of genes with 2-fold down-regulation in the given category
f: number of genes total in the given category

3.3.3 Identification of Whi5p as a new component in Swi4/6 complex

Affinity purification approach was employed to purify Swi6p-associated proteins. Swi6p was known to interact with Swi4p and Mbp1p that were observed in the purification (Figure 3.9A). Additionally, Whi5p and Ykr018cp were found in the purification (Figure 3.9A). To confirm the association of Whi5p with Swi6p, reciprocal affinity purification by using Whi5-TAP as the bait was performed. Swi6p, Swi4p, and Ykr018cp were observed in the purified sample, except Mbp1p, suggesting that Whi5p may not be in MBF complex (Figure 3.9B). Based on these data, a model was proposed for SBF and MBF complexes, with Whi5p and Ykr018cp interacting with SBF complex (Illustration 3.3). Mutant with deletion of *WHI5* exhibited a small cell size and shortened doubling time (Jorgensen et al., 2002). The exact role of *WHI5* was not clear by then. During the progress of my work, Whi5p was characterized to be a negative regulator for G1 to S phase transition by directly interacting with SBF and MBF complexes (Costanzo et al., 2004; de Bruin et al., 2004). However, Ykr018cp is still not characterized.

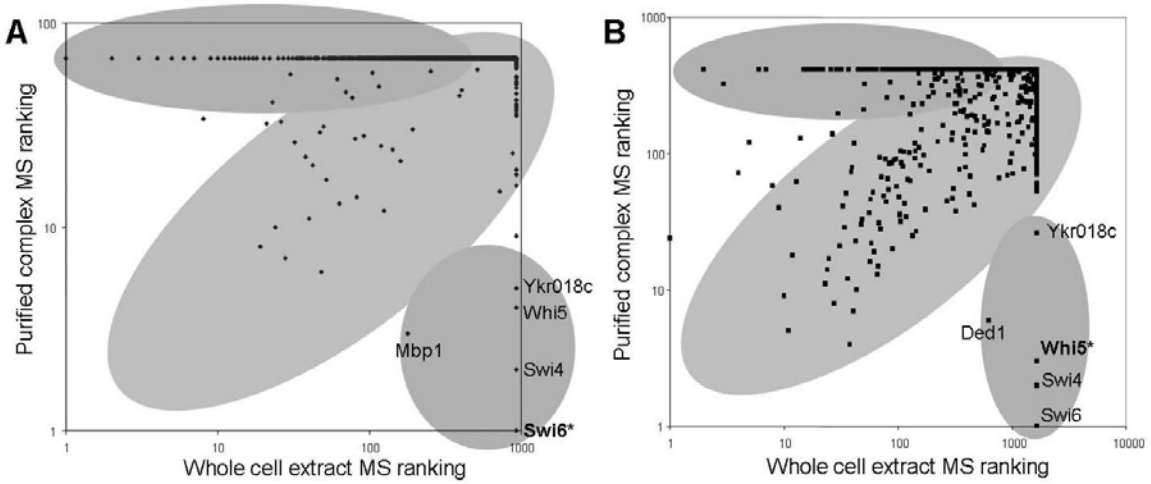


Figure 3.9 TAP purifications by using Swi6-TAP or Whi5-TAP as baits.
Specific purified proteins in purified samples locate in the bottom-right area of the plot, showing high enrichment in the purified samples. The proteins located on the diagonal area are mostly contaminants in the purifications. Swi6-TAP (A) and Whi5-TAP (B) were used as the baits (labeled with star) in the purifications.

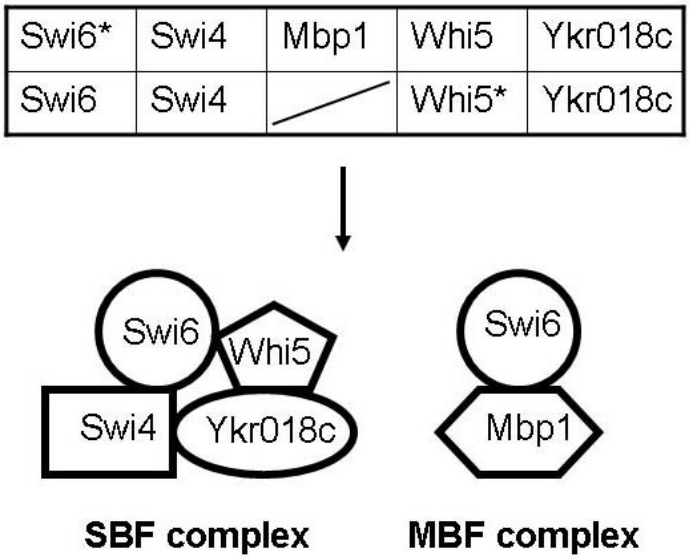


Illustration 3.3 A model for SBF and MBF complexes.
Based on the purification data from Figure 3.9, SBF complex (Swi6p and Swi4p) will include two more components: Whi5p and Ykr018cp, while MBF complex contains Swi6p and Mbp1p. Star indicates the bait protein used in the purification.

3.3.4 Improvement of complex affinity purification

During this work, I found that TAP purification requires large amount of culture. For example, 8 L or 10 L are required to purify the Hap2/3/5-Gln3p complex. This is partly attributed to the low abundance of the transcription factors, but also the inefficiency of TAP purifications. It has been shown that the efficiency of the first-step purification and the second-step purification are 80% and 50%, respectively (Rigaut et al., 1999). Therefore, the two-step purification efficiency is only 40%. To reduce the large amount of culture and increase the purification efficiency, I employed a simple and effective method to purify protein complexes (Illustration 3.4). One-step affinity purification was used and the bound protein complexes and the non-specific bound proteins were eluted by trypsin digestion buffer plus heat, or acidic buffer. The wild type control cells were also used for purification as the control for the non-specific bound proteins. The eluted proteins were digested to peptides that were directly analyzed by MudPIT (multi-dimensional protein identification technology). By comparing identified proteins from the control sample with those from the TAP strain sample, the proteins enriched in the TAP strain sample can be identified. One such example is to utilize Swi6-TAP as the bait to perform one-step purification. By using trypsin digestion buffer plus heat to denature and elute Swi6p-interacting proteins, Swi4p, Mbp1, and Whi5p were identified (Figure 3.10A), whereas Swi6p, Swi4p, and Whi5p were identified by using acidic buffer to elute the proteins bound to IgG beads (Figure 3.10B). By either procedure, only 1.5 L culture was required, compared to 8 L culture for two-step affinity purifications (Figure 3.10). There were many non-specific bound proteins in one-step affinity purification. However, they could be easily distinguished from the true positives by comparing with the control purification (Figure 3.10). This method should also be useful for other tags besides TAP tag.

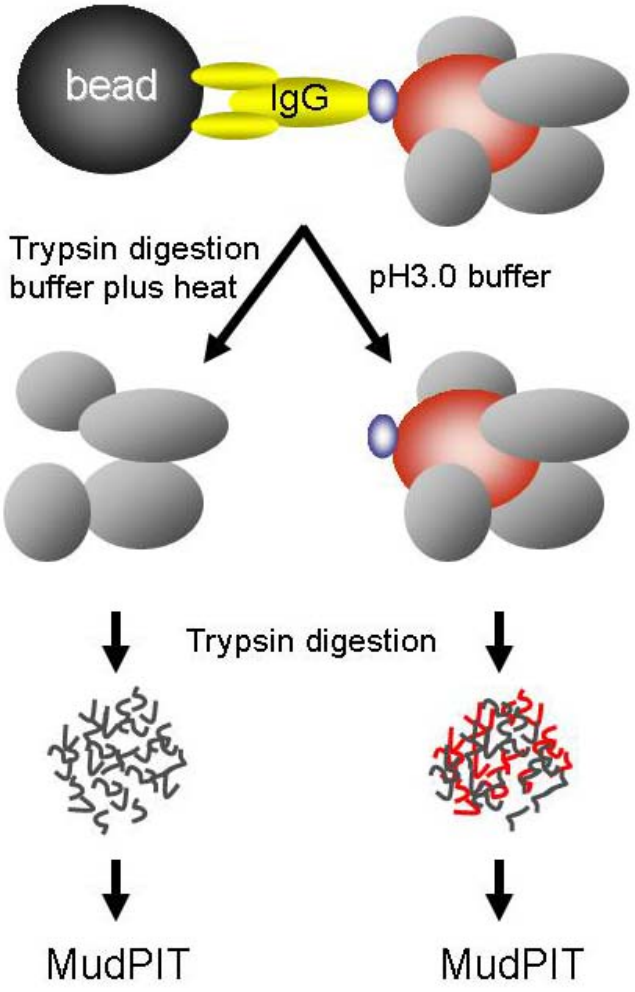


Illustration 3.4 One-step affinity purification coupled with MudPIT.
IgG beads were used to incubate with the samples prepared from wild type control strain and TAP-tagged strain. After extensive washing, the proteins bound to IgG beads were eluted by two alternative methods: trypsin digestion buffer (100 mM Tris pH 8.0) plus heat; acidic buffer (Glycine-HCl buffer, pH 3.0). The eluted proteins were digested with trypsin and analyzed with MudPIT.

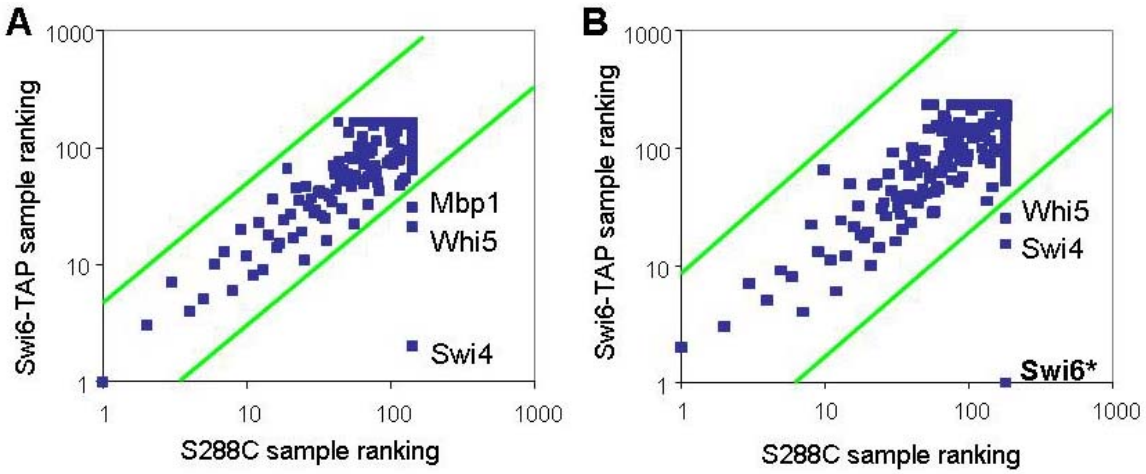


Figure 3.10 One-step affinity purification of Swi6p-associated proteins.
Affinity purification by using Swi6-TAP as the bait was performed as Illustration 3.4. The MS rankings for purified proteins from Swi6-TAP strain (y axis) were plotted against the MS ranking for purified proteins from the wild type control strain S288C (no TAP tag) (x axis). The proteins along the diagonal (within the green lines) existed in both specific (TAP-tagged strain) and non-specific purifications (strain without TAP tag), and were the contaminants in the purifications. By using trypsin digestion buffer plus heat method to elute the bound proteins from the beads (A), Swi4p, Whi5p, and Mbp1p were identified but not the bait protein, Swi6p, possibly due to the strong binding of Swi6-TAP to IgG beads. By using acidic buffer to disrupt the interaction between Swi6-TAP and IgG beads, Swi4p, Whi5p, and the bait protein Swi6p (labeled with star) were identified, but not Mbp1p, possibly due to the overwhelming contaminants that could mask identification of low abundant proteins in mass spectrometry analysis.

3.4 DISCUSSION

Several transcription factors were genomically tagged by TAP tag at their carboxyl-terminus in yeast cells, and three transcription factor complexes were identified: histone deacetylase complex B (HDB), Hap2/3/5 complex, and Swi6-containing complexes. Additionally, a simple and effective one-step protein purification and identification method was developed.

Affinity purification of HDB by using Rpd3-TAP as the bait identified several novel components in addition to the known proteins in HDB (Figure 3.1). Dep1p has been suggested to belong to a class of transcription factors including Rpd1p, Sin3p, Rpd3p, and Spt10p by analyzing the mutant phenotypes (Lamping et al., 1994). Recently, two distinct Rpd3-containing complexes were purified: Rpd3L and Rpd3S (Carrozza et al., 2005). Three components in my purified complex: Dep1, Ybr095cp (Rxt2p), and Ymr263cp (Sap30p) belong to the Rpd3L complex, whereas Ume1p and Ymr075w (Rco1p) are components of Rpd3S complex. However, Fun19p is still function unknown. Based on my affinity purification data, Fun19p is likely to be involved in histone deacetylation.

Hap2/3/5 complex is well characterized as a respiratory regulator (McNabb et al., 1995). Interestingly, Gln3p was consistently co-purified with this complex (Figure 3.2). Co-immunoprecipitation assay confirmed that low level of Gln3p co-purified with Hap2/3/5 complex (Figure 3.4). Additionally, affinity purification with EtBr also identified Gln3p, suggesting a DNA-independent interaction (Figure 3.3). However, affinity purification by using Gln3-TAP as the bait failed to purify Hap2/3/5 complex (Figure 3.2). Instead, it purified the known interacting components in the cytoplasm (Figure 3.2). Although reciprocal purification can confirm the interaction, it is possible that two proteins can not be reciprocally purified due to the tag on one protein that will

interfere with the interaction. Another possibility is that low abundant Gln3p exists in the nucleus and interacts with Hap2/3/5 complex, whereas most Gln3p are in the cytoplasm under rich medium condition. Purification by using Gln3-TAP as the bait identified the abundant complex in the cytoplasm instead of the low abundant nuclear complex.

The function of Hap2/3/5 complex depends on each component in the complex. Deletion of any one of the components in Hap2/3/5 complex leads to the incompetence of its function (McNabb et al., 1995). A strain with deletion of both *HAP3* and *GLN3* was constructed to disrupt the functions of Hap2/3/5 complex and Gln3p in order to examine genetic interactions between Hap2/3/5 complex and Gln3p. *hap3gln3Δ* exhibited slight growth defect compared to wild type strain and single deletion mutant, and no obvious morphological defects (Figure 3.5).

Gln3p is a transcription factor that regulates nitrogen assimilation (Courchesne and Magasanik, 1988), while Hap2/3/5 complex regulates genes in response to carbon source (Gancedo, 1998). The physical interaction between Hap2/3/5 complex and Gln3p suggested a crosstalk between nitrogen metabolism and carbon metabolism. Indeed, *GDH1* is a known gene regulated by both Gln3p and Hap2/3/5 complex as well as involved in the crosstalk between nitrogen and carbon metabolism (Riego et al., 2002). To elucidate the functional relationship between Hap2/3/5 complex and Gln3p, gene expression analyses by DNA microarrays were performed for single deletion and double deletion mutants under rich medium condition or rapamycin treatment condition. As expected, many metabolism genes were regulated by Hap2/3/5 complex and Gln3p, from which a fraction of genes were regulated by both transcription factors. This pattern was also observed in a large-scale genome-wide study to identify the transcription factor binding sites under different conditions (Harbison et al., 2004). Upon rapamycin treatment, promoter regions of 20 genes were found to be bound by both Hap2p and

Gln3p with significant cutoff ($p < 0.001$), promoters of 22 genes were only bound by Hap2p, and promoters of 47 genes were only bound by Gln3p (Harbison et al., 2004). However, only 2 genes (*DAL2* and *DAL3*) out of the 20 genes overlap with my gene expression data; 7 genes out of Hap2p-bound 22 genes and 13 genes out of Gln3p-bound 47 genes overlap with my gene expression data. The small overlap between gene expression data and genome-wide binding data might be partly attributed to the different techniques. Gene expression data describe directly and indirectly regulated genes, and provide information about positive and negative regulation. Whereas genome-wide location analysis identifies directly bound regions by a transcription factor, but does not provide information about activation state of the transcription factor. False positive and false negative identifications from both methods might also contribute to the discrepancy of the small overlap. For example, *GDH1* is known to be regulated by both Hap2/3/5 complex and Gln3p. Genome-wide location analysis failed to identify *GDH1* as a Gln3p-bound gene. All the overlapped genes except *ECM13* exhibited down-regulation in the deletion strains, suggesting that Hap2/3/5 complex and Gln3p function as activators in regulating gene expression. The up-regulated genes in deletion mutants might be secondary effects. Nonetheless, both gene expression studies and genome-wide location analyses provide information about the embedded functional relationship between Hap2/3/5 complex and Gln3p.

Affinity purification by TAP tag is a powerful technique, and recently it was employed by two groups to purify protein complexes for the whole yeast proteome, revealing the proteome-wide protein-protein interactions (Gavin et al., 2006; Krogan et al., 2006). It was also successfully applied to other organisms such as plants and human (Brajenovic et al., 2004; Forler et al., 2003; Knuesel et al., 2003; Rivas et al., 2002). The

data provided by this method will be invaluable for dissecting protein functions in functional genomic studies.

Chapter 4: High-throughput Mapping Native Protein Complexes by Density Gradient Coupled with Two-dimensional Liquid Chromatography Mass Spectrometry

4.1 INTRODUCTION

In cells, many proteins interact with each other to form protein complexes as the functional entities. For example, in yeast, the 40S ribosomal subunit consists of 32 ribosomal proteins (Planta and Mager, 1998), and the 20S proteasome is composed of 14 different proteins with 2 copies of each protein (Heinemeyer et al., 1994). Thus, protein complex mapping is necessary not only for identifying protein interaction partners but also for deciphering functional relationships among proteins. Unlike yeast two-hybrid method that detects binary interactions, protein complex mapping obtains protein physical association information that yeast two-hybrid method might not detect. For example, Whi5p is a component of SBF complex (Swi6p-Swi4p), and Whi5p can not interact with Swi6p without Swi4p, or vice versa (de Bruin et al., 2004). Protein complex mapping would detect such relationship and provide invaluable information to reveal protein functions.

Currently, affinity purification coupled with mass spectrometry identification method has been successfully applied to map protein complexes in a proteome-wide scale in yeast (Gavin et al., 2006; Krogan et al., 2006). This approach has many advantages as discussed in chapter 3. However, this method usually involves a tag that is used for affinity purification. In some cases, the tag will disrupt protein-protein interactions and affect their functions. For example, the carboxyl-terminal GFP-tagged Rpl25p or Rpl35p reduced the binding of Arx1p to the 60S ribosomal subunit (Hung and Johnson, 2006). Additionally, in order to maintain physiological interactions, the tag was usually introduced at the carboxyl-terminus of the gene in its genomic locus, which has been

done on a genomic-scale in yeast and E. coli (Butland et al., 2005; Ghaemmaghami et al., 2003). However, in plant and mammalian cells, genetic manipulation is not as easy as in yeast cells. Some cell lines such as mouse embryonic stem cells and chicken DT40 pre-B cells have been used for tagging genes at their carboxyl-terminus (Chen et al., 2006). But it would be a daunting task to tag the whole proteome due to technical difficulties and the labor involved. Therefore, new strategies for protein complex mapping are needed.

Different complexes have different sizes, shapes, and charges due to the composition of different proteins and their stoichiometry in each complex. Those properties can be utilized to separate and purify the different complexes by using biochemical techniques such as density gradient ultracentrifugation, gel filtration chromatography, ion-exchange chromatography, native gel electrophoresis, and isoelectric focusing. For example, the RNA polymerase II complex was purified by chromatography (Young, 1991). The 20S proteasome was purified by ion-exchange and gel filtration chromatography followed by sucrose density gradient (Arrigo et al., 1988). Spliceosomes can be isolated by gel filtration chromatography combined with affinity purification (Reed, 1990). The mitochondrial respiratory complexes has been purified by both native gel electrophoresis (Schagger and Pfeiffer, 2000) and sucrose density gradient ultracentrifugation (Dudkina et al., 2005). More recently, blue native polyacrylamide gel electrophoresis (BN-PAGE) was used to separate protein complexes from whole cell lysates (Camacho-Carvajal et al., 2004). All those techniques provide a generic approach to separate protein complexes from any cell type under native conditions.

Biochemical purification of protein complexes has a long history, but in the past, it is a difficult task to reveal the identities of the components of the complex. One approach that was applied to study the components of the RNA polymerase II complex is to use antibodies against the purified proteins to probe the phage expression library to

isolate the genes for specific components in the complex (Young and Davis, 1983). Another way is to obtain a short peptide sequence information by Edman degradation method, followed by screening genomic DNA library based on all the codon choices for this peptide (Hewick et al., 1981; Schmidt et al., 1989). Those methods are time-consuming and labor-intensive. The development of mass spectrometry technologies to analyze proteins and peptides revolutionized protein identifications (Fenn et al., 1989; Karas and Hillenkamp, 1988). By combining with database searching, mass spectrometry-based protein identifications greatly speeded up protein identifications (Henzel et al., 1993). In one example, six new components were identified from the purified U1 small nuclear ribonucleoprotein complex by mass spectrometry-based approach (Neubauer et al., 1997). In the aid of computational algorithm for database searching and multidimensional liquid chromatography separation of peptide mixtures, the components of protein complexes such as ribosomes could be analyzed directly by mass spectrometer (Link et al., 1999). This approach was later developed as the multidimensional protein identification technology (MudPIT) (Washburn et al., 2001), also referred to shotgun proteomics (Wolters et al., 2001). Currently, mass spectrometry has become the major approach for protein identifications.

Mass spectrometry could also be utilized to quantify protein abundances in the sample. One method is to use isotope-labeled peptides as references to relatively or absolutely quantify proteins in the sample, such as isotope-coded affinity tag technology (ICAT) (Gygi et al., 1999), isotope-coded protein label (ICPL) (Schmidt et al., 2005), isobaric tag for relative and absolute quantitation (iTRAQ) (Ross et al., 2004), stable isotope labeling by amino acid in cell culture (SILAC) (Ong et al., 2002), and absolute quantification by isotope labeled internal control peptides (AQUA) (Gerber et al., 2003). Another one is a label-free approach, relatively simple for both experimentation and data

analysis when compared to the methods based on isotope-labeling. Either peptide peak intensities or spectral counts can be utilized to quantify protein abundances (Gao et al., 2003; Liu et al., 2004; Wang et al., 2003). Based on spectral counts, two absolute protein quantification indexes were derived: exponentially modified protein abundance index (emPAI) (Ishihama et al., 2005) and absolute protein expression (APEX) index (Lu et al., 2007). A comparison study has shown that peak intensities and spectral counts methods agreed to each other in general with more sensitivity for spectral count measurements and more accuracy for peak intensity measurements (Old et al., 2005).

In an effort to identify protein components of the centrosome in human cells, by combining centrosome purification, sucrose density gradient, and mass spectrometry to analyze the proteins in each fraction of the sucrose density gradient, 23 novel proteins were identified to be the components of the human centrosome (Andersen et al., 2003). By using mass spectrometry to identify and quantify the proteins in different organelles that were separated by sucrose density gradient, over thousand proteins were assigned to ten cellular organelles from mouse liver (Foster et al., 2006). Both studies were based on protein correlation profiling by combing traditional separation techniques with state-of-the-art protein identifications and quantifications by mass spectrometry. This method could possibly be applied in general to identify protein complexes in the cell. After separating protein complexes with biochemical techniques, mass spectrometry can be utilized to analyze each protein's abundance across all the fractions collected from the biochemical separation to obtain the abundance profile for each protein. Proteins with similar abundance profiles will likely to be in the same complex. In a preliminary study, sucrose density gradient was applied to separate large protein complexes, and the proteins across the fractions from the sucrose density gradient were analyzed by mass spectrometry to obtain the protein abundance profiles (Illustration 4.1).

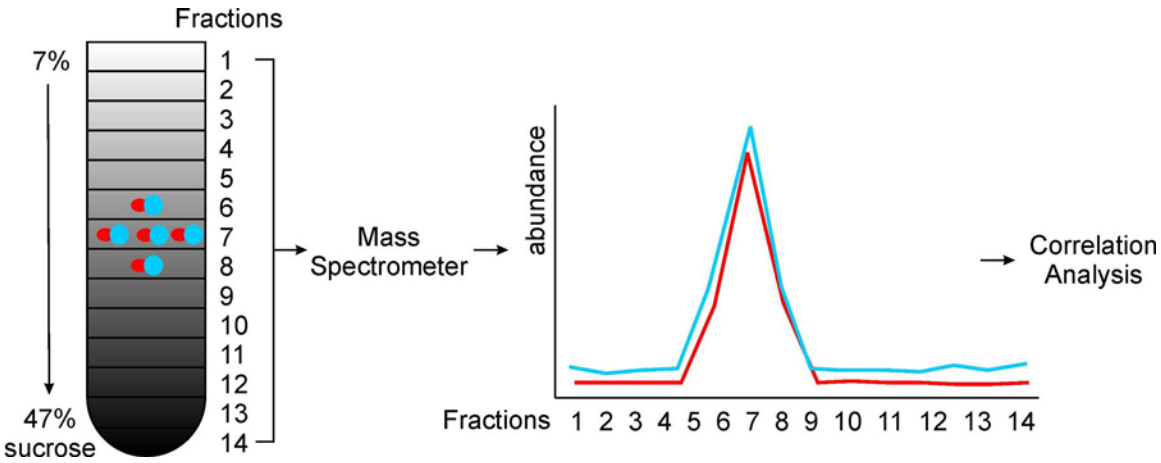


Illustration 4.1 A schematic diagram to identify co-sedimented proteins.
Whole cell lysate was loaded onto a sucrose density gradient and different complexes were separated by their size and shape. Fractions were collected and proteins in each fraction were analyzed and quantified by a shotgun proteomic approach. The protein abundance profiles can be further analyzed by correlation analysis to identify the proteins with similar profiles that are potentially in the same complex.

4.2 MATERIALS AND METHODS

4.2.1 Sucrose density gradient and mass spectrometry of yeast total cell lysate

4.2.1.1 Yeast culture and sucrose density gradient

The wild type strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was cultured to early logarithm phase in YPD at 30°C. 200 ug/ml cycloheximide (Sigma) was added to each culture. Cultures were immediately cooled with ice and all subsequent steps were performed on ice or at 4°C. Cells were collected by centrifugation at 3000g for 5 minutes and washed once with lysis buffer (20 mM Tris pH 7.4, 20 mM KCl, 5 mM $MgCl_2$, 200 ug/ml cycloheximide, 12 mM β-mercaptoethanol). The cell pellet was resuspended in the lysis buffer with protease inhibitors (2 ug/ml leupeptin, 2 ug/ml aprotinin, 1 ug/ml bestatin, 1 ug/ml pepstatin A, 1mM PMSF) and lysed by vortexing with glass beads. Crude lysates were clarified by centrifuging at 15,000g for 10 minutes. Fifteen OD_{260} units of the supernatant were loaded onto a continuous 7 to 47% sucrose gradient. After centrifuging for 2.5-hr at 40,000 rpm in a Beckman SW40 rotor, the sucrose gradient was fractionated (ISCO fractionator) and fractions were collected. Proteins from each fraction were precipitated with 10% cold trichloroacetic acid, washed with 100% cold acetone, and dried in the air.

4.2.1.2 Mass spectrometry

Protein pellets were resuspended in 100 mM Tris buffer (pH 8.0), and digested with proteomic grade trypsin (Sigma) for 24 hrs at 37°C. Each digested sample was diluted with 94.9% water, 5% acetonitrile, 0.1% formic acid, filtered with a Microcon 10 filter (Millipore). The peptide mixture was first separated by a strong cation exchange column with a four-step salt elution: 5, 20, 60, and 900 mM ammonium chloride. Each salt eluate was loaded directly onto a reverse phase C18 column. After washing off the

salts from the C18 column, a 125-min continuous gradient from 5%-50% acetonitrile (ACN) was used to elute off the peptides that were subsequently analyzed online with an electrospray ionization (ESI) ion trap mass spectrometer (ThermoFinnigan DecaXPplus). Each sample was analyzed three times at different mass/charge (m/z) range (300-650, 650-900, and 900-1500) for parent ions. For each parent ion mass spectrum, the top five spectra were selected for fragmentation by collision induced dissociation (CID) with helium gas to produce MS/MS spectra. The MS/MS spectra from three runs of different m/z range were combined and used to search against a database of Saccharomyces cerevisiae protein sequences by SEQUEST algorithm (Bioworks 3.2, Thermo). Proteins were identified at a 5% false detection rate by using PeptideProphet and ProteinProphet (Nesvizhskii et al., 2003), and quantified by spectral counts (total number of MS/MS spectra for each protein). For each sucrose gradient fraction, the frequency of each protein in each fraction was calculated as $f_{ij} = (n_{ij}/N_j)*10000$, where f_{ij} is the relative frequency for protein i in fraction j, n_{ij} is the spectral count for protein i in fraction j, and N_j is the total number of observed spectra in fraction j. The normalized frequency of each protein was used to plot the protein's abundance profile across the sucrose density gradient. The frequency of each protein was further normalized across the gradient: $f_{ij}' = (f_{ij}/\sum_J f_{ij})*1000$, which was used for hierarchical clustering (Eisen et al., 1998).

4.2.1.3 Calculation of the center of mass for each protein in sucrose density gradient

Each protein is distributed across all the density fractions. The center of mass of this distribution reflects the fraction in which this protein is most enriched, calculated as $C_i = \sum_J (f_{ij}*j)/\sum_J f_{ij}$, which was used for hierarchical clustering.

4.2.1.4 Calculation of the protein profile correlation score

A chi-squared goodness of fit test was used to test whether the observed distribution for the categorical variable (f_{ij} that reflects the protein's abundance level) across the density fractions differs from the hypothetical distribution that was calculated based on a group of similar protein profiles across the gradient fractions. The statistic software STATA version 8 and the **csgof** module developed by UCLA (http://www.ats.ucla.edu/stat/stata/whatstat/whatstat.htm) were used to perform the analyses. The calculated chi-square values were used for clustering analysis.

4.2.2 Sucrose density gradient and mass spectrometry of Hela cell lysate

4.2.2.1 Cell culture and sucrose density gradient

Hela S3 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37 °C with 5% CO2. At about 80% confluency, cells were treated with 100 μg/ml emetine for ten minutes and harvested by scraping. Cells were centrifuged at 500g for ten minutes, washed three times with cold PBS buffer, and resuspended in five packed cell volumes of cold lysis buffer (10mM Tris pH7.4, 20mM KCl, 5mM MgCl2). After swelling on ice for ten minutes, cells were centrifuged at 500g for ten minutes and resuspended in one packed cell volume cold lysis buffer supplemented with 1X protease inhibitor cocktail (Roche) and 100μg/ml emetine. After lysing the cells with a dounce homogenizer, nuclei were collected by centrifuging 1,000g for ten minutes. The supernatant was centrifuged at 15,000g for ten minutes to obtain the cytosolic fraction. Nuclei were suspended in lysis buffer and lysed by sonication, collecting the clarified supernatant after centrifugation at 15,000g for ten minutes. Sucrose density gradients for cytosolic and nuclear fractions and protein precipitations were performed as in section 4.2.1.1.

4.2.2.2 Mass spectrometry

Protein pellets were suspended in 100 mM pH 8.0 Tris buffer and digested with proteomic grade trypsin (Sigma). Tryptic peptides were loaded onto a reverse phase C18 column and washed with 94.9% water, 5% acetonitrile, 0.1% formic acid. Peptides were separated and eluted with a 240-min gradient from 5% to 40% acetonitrile and analyzed online with electrospray ionization ion trap mass spectrometry using an LTQ-Orbitrap hybrid mass spectrometer (Thermo Electron) using data-dependent precursor ion selection. Each parent ion mass spectrum (MS) was analyzed at high-resolution (100,000) with the Orbitrap; the top seven MS peaks were fragmented by helium collision-induced dissociation (CID) at 35 eV, analyzing the resulting MS/MS spectra with the LTQ. Approximately 35,000 MS/MS spectra were collected per fraction. Spectra were searched against the set of NCBI human protein sequences using TurboSequest (Bioworks v. 3.2, Thermo Electron). Proteins from each fraction were identified at a 5% false positive rate using Peptide/ProteinProphet (Keller et al., 2002; Nesvizhskii et al., 2003). The spectral count (number of total observations of MS/MS spectra from a given protein in a given fraction) was used as an estimate of protein abundance, normalizing the spectral count of a protein by the sum of spectral counts for all proteins identified in that fraction.

4.3 RESULTS

4.3.1 Sucrose density gradient and mass spectrometry of yeast total cell lysate

4.3.1.1 Separation of different sizes of ribosomal particles by sucrose density gradient

In eukaryotic cells, the pre-ribosome is first assembled in nucleolus as a 90S complex that is subsequently processed into 66S and 43S pre-ribosomes, which continue to mature into 60S and 40S ribosomal subunits along the path from the nucleolus to the cytoplasm (Tschochner and Hurt, 2003). During translation, ribosomal subunits are

loaded onto mRNAs to form polysomes with different number of 80S ribosome units. Due to large contents of rRNAs in the ribosome particles, the major ribosomal particles (40S, 60S, 80S, and polysomes) were clearly visible by monitoring the absorbance at 254 nm in a sucrose density gradient (Figure 4.1). The peaks for 90S, 43S, and 66S were not identifiable from the polysome profile due to the low abundance of those particles and their sizes close to the abundant 80S, 40S, and 66S particles.

Mass spectrometry analyses of the proteins across 14 fractions from the sucrose density gradient identified 1,023 unique proteins in total. The identified proteins clustered into four major groups (Figure 4.2A). One group was highly enriched for metabolic enzymes and they primarily distributed in the low density fractions in the sucrose density gradient. A group of proteins located in the fractions corresponding to the 60S subunits, many of which are factors involved in 60S subunit biogenesis. Similarly, many 40S biogenesis factors clustered together and mainly located in the 40S fractions. Most of the ribosomal proteins clustered together and distributed in the high density fractions corresponding to the polysomes. Distributions of representatives for each group were shown in Figure 4.2B-E.

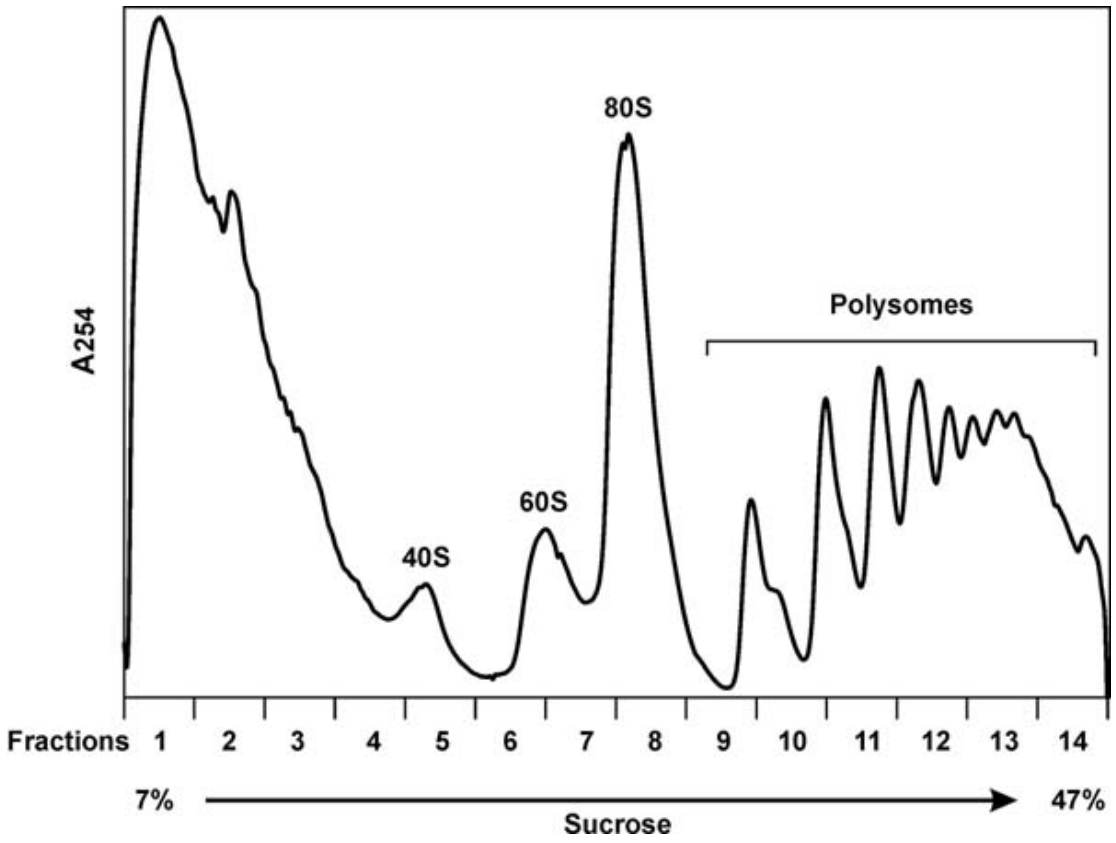


Figure 4.1 Sucrose density gradient separation of yeast whole cell extracts.
Ribosomal particles were separated in a continuous 7-47% sucrose density gradient and the fractions were collected with fraction 1 corresponding to the lowest density fraction and fraction 14 corresponding to the highest density fraction. The Y-axis indicates the absorbance at 254 nm wavelength that measures the quantities of the nucleic acids in the sample.

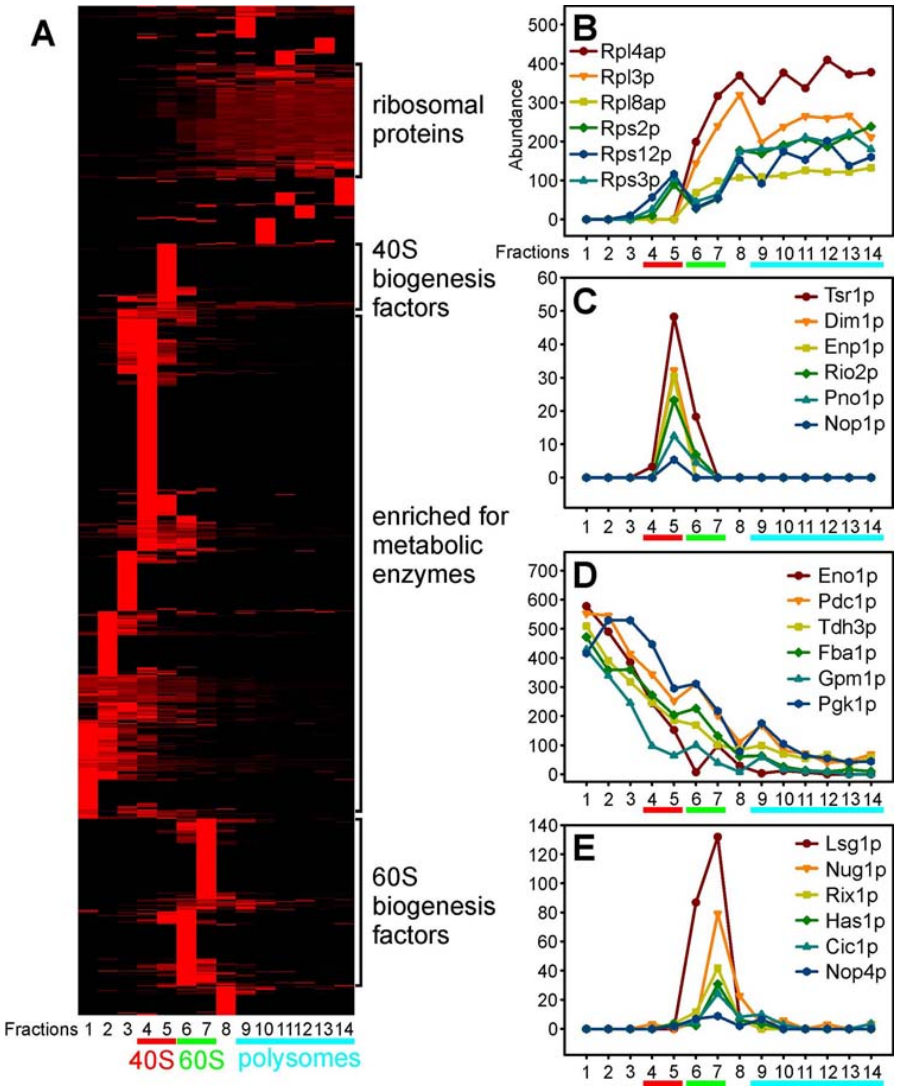


Figure 4.2 Mass spectrometry identification of proteins in fractions of sucrose density gradient for yeast whole cell lysate.

(A) A clustergram of identified proteins from fractions of the sucrose density gradient. Each row represents each gene identified from mass spectrometry. Each column represents each fraction collected from sucrose density gradient, with fraction 1 as the lowest density fraction and fraction 14 as the highest density fraction. Four distinct clusters have enriched ribosomal proteins, 40S biogenesis factors, metabolic enzymes, and 60S biogenesis factors. Fractions 4 and 5 correspond to 40S peak in polysome profile; fractions 6 and 7 correspond to 60S peak in polysome profile; and fractions 9-14 correspond to polysomes in polysome profile.

(B-E) Representative protein abundance profiles based on mass spectrometry identifications and quantitations for ribosomal proteins (B), 40S biogenesis factors (C), metabolic enzymes (D), and 60S biogenesis factors (E).

4.3.1.2 Assign each identified protein into a fraction based on its center of mass

To identify the most enriched fraction for each protein based on its distribution across the sucrose density gradient, the center of mass for each protein was calculated as described in the methods. The values of the center of mass were used for clustering and four distinct groups were identifiable (Figure 4.3). Most metabolic proteins were enriched in the first four fractions. The 40S subunit biogenesis factors were enriched in fraction 5 and the 60S subunit biogenesis factors were enriched in fraction 6 and 7. Additionally, some ribosome biogenesis factors resided in fractions 8 and 9, which might indicate that these proteins were in the 90S pre-ribosomal complex. The ribosome proteins were enriched in fractions 10 and 11. The membrane proteins, mitochondria proteins, and vacuole proteins were enriched in the high density fractions because of their sedimentation properties. This approach is good for those proteins with an obvious unimodal distribution, and it might not be appropriate for the proteins with a bimodal distribution.

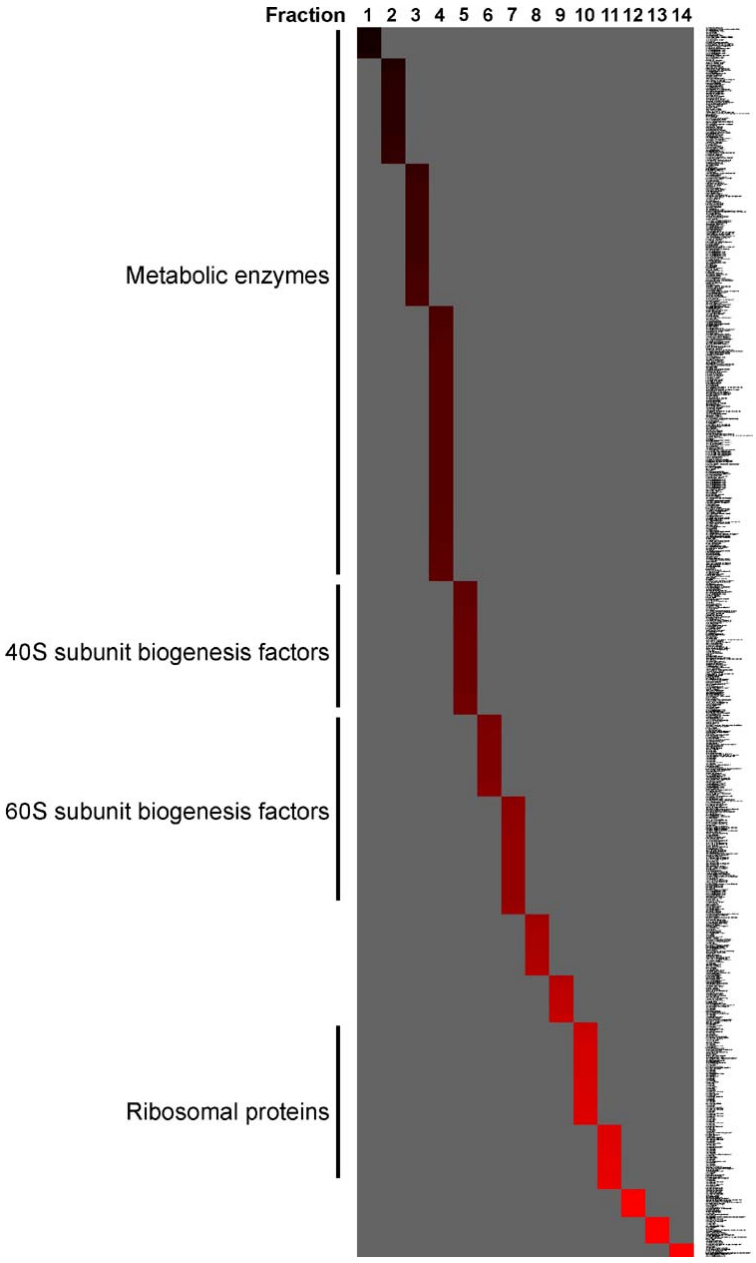


Figure 4.3 Clustering of the proteins based on the center of mass for each protein across the density gradient.
Each row represents each gene identified from mass spectrometry. Each column represents each fraction collected from sucrose density gradient, with fraction 1 as the lowest density fraction and fraction 14 as the highest density fraction. Most metabolic enzymes had a density point within the first four low density fractions; the 40S subunit biogenesis factors' density point was in fraction 5 and the 60S subunit biogenesis factors' density point was in fraction 6 and 7; the ribosomal proteins' density point was in fraction 10 and 11.

4.3.1.3 Protein profile correlation analysis to assign each protein into different functional groups

Different proteins from different functional categories (the metabolic enzymes, the 40S subunits, the 60S subunits, and the ribosome proteins) had different distribution patterns across all the fractions (Figure 4.2B-E), which could be used as the references to find other proteins with similar distribution patterns. Proteins with similar distribution patterns would be likely in the same complex and in the same functional category. First, some known proteins from the four functional categories were selected (Table 4.1). For each functional category, the average abundance level of the proteins was calculated for each fraction to establish an averaged abundance distribution pattern as the reference (Figure 4.4A-D). Then the protein profile correlation scores were calculated for each protein by comparing the protein's distribution pattern to the four references using chi-square goodness of fit test. Each protein would have four scores that reflected the likelihood of this protein within each specific functional category. After clustering analysis, the proteins belonging to the same functional category would be grouped together (Figure 4.5). Most metabolic enzymes clustered in the free fraction (the low density fractions); the known 40S and 60S biogenesis factors clustered in the 40S and 60S fractions respectively; the ribosome proteins were in the polysome fractions corresponding to the high density fractions.

Table 4.1 The reference proteins of the four functional categories.

Functional category	ORF	Gene	Biological process
metabolic enzymes (free fractions)	YHR174W	ENO2	glycolysis
	YGR254W	ENO1	glycolysis
	YLR044C	PDC1	pyruvate metabolism
	YGR192C	TDH3	glycolysis
	YKL060C	FBA1	glycolysis
	YKL152C	GPM1	glycolysis
	YCR012W	PGK1	glycolysis
	YLL024C	SSA2	response to stress
	YAL038W	CDC19	glycolysis
	YPL106C	SSE1	telomere maintenance
	YOL086C	ADH1	fermentation
	YMR186W	HSC82	telomere maintenance
	YJR045C	SSC1	protein folding
	YIL053W	RHR2	response to osmotic stress
	YER043C	SAH1	methionine metabolism
40S biogenesis (40S fractions)	YDL060W	TSR1	rRNA processing
	YPL266W	DIM1	rRNA modification
	YBR247C	ENP1	35S primary transcript processing
	YNL207W	RIO2	processing of 20S pre-rRNA
	YOR145C	PNO1	35S primary transcript processing
	YOR056C	NOB1	processing of 20S pre-rRNA
	YOR119C	RIO1	processing of 20S pre-rRNA
	YNL308C	KRI1	ribosome biogenesis
60S biogenesis (60S fractions)	YER036C	ARB1	ribosome biogenesis
	YER006W	NUG1	rRNA processing
	YLR106C	MDN1	rRNA processing
	YJL050W	MTR4	35S primary transcript processing
	YMR290C	HAS1	rRNA processing
	YHR197W	RIX1	35S primary transcript processing
	YHR052W	CIC1	ribosomal large subunit biogenesis
	YPL043W	NOP4	rRNA processing
ribosomes (polysome fractions)	YBR031W	RPL4A	translation
	YDR012W	RPL4B	translation
	YOR063W	RPL3	translation
	YML063W	RPS1B	translation
	YGL123W	RPS2	translation
	YLR441C	RPS1A	translation
	YNL178W	RPS3	translation
	YOR369C	RPS12	translation

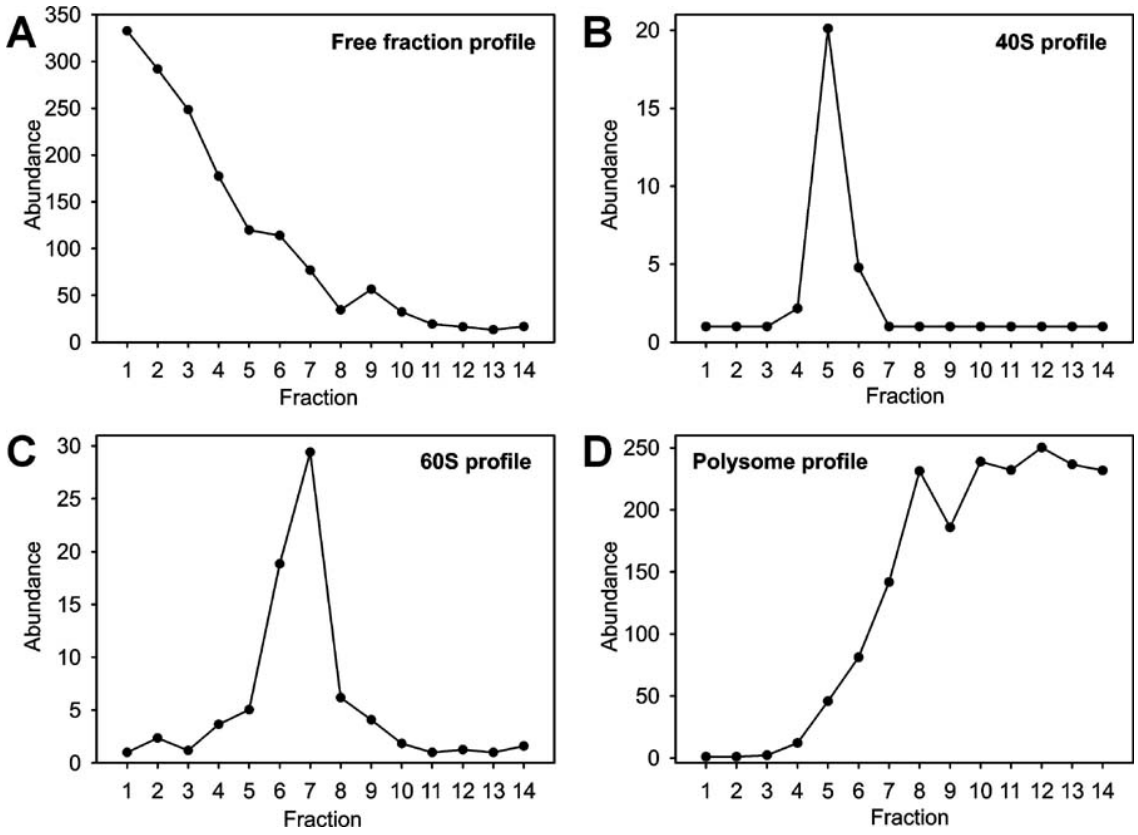


Figure 4.4 The reference protein profiles along the density gradient fractions.
(A) The averaged protein profile of the known metabolic enzymes listed in Table 4.1. The metabolic enzymes were enriched in the low density fractions (the free fraction).
(B) and (C) showed the averaged protein profiles of the known 40S and 60S biogenesis factors respectively with 40S biogenesis factors enriched in fraction 5 and 60S biogenesis factors enriched in fractions 6 and 7.
(D) The averaged profile of the ribosomal proteins listed in Table 4.1 that were enriched in polysome fractions.

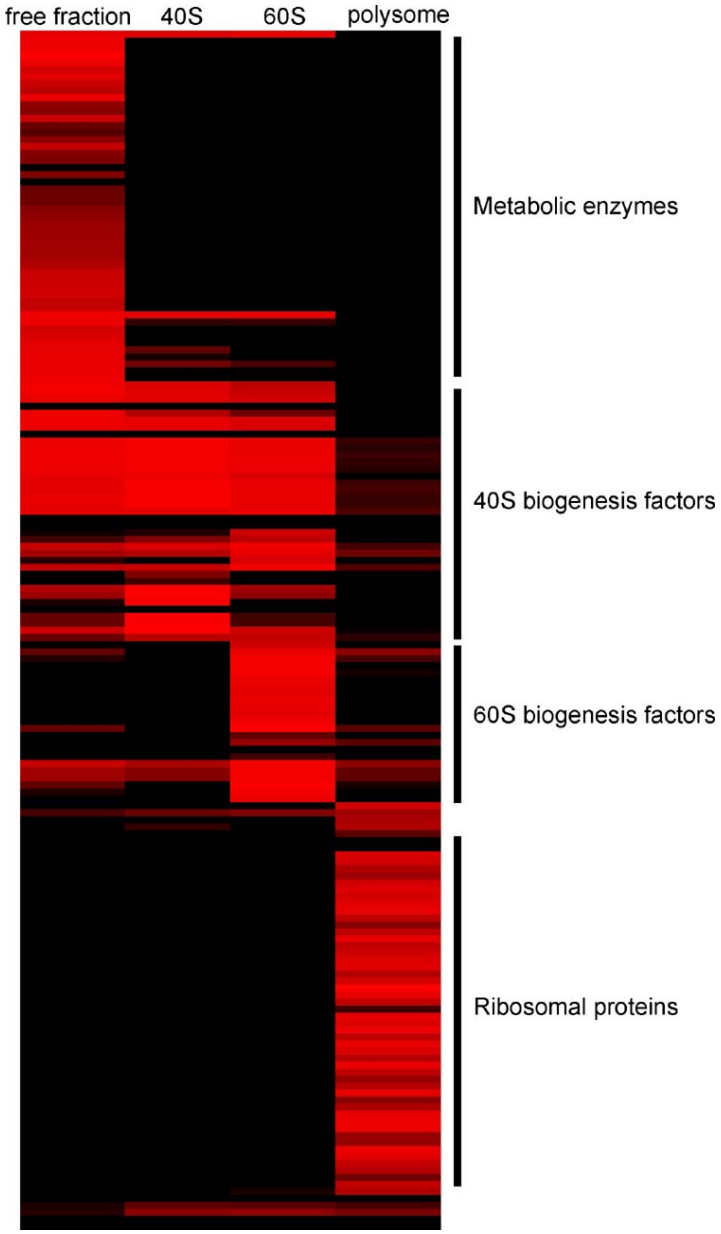


Figure 4.5 Clustering of the protein profile correlation scores.
Each protein's correlation scores with respect to the four characteristic profiles (Figure 4.4) were calculated and clustered. Each row represents a protein out of 167 proteins in this clustergram. Each column represents one of the four functional categories. Most metabolic enzymes correlated with the free fraction profile in Figure 4.4A. The 40S and 60S biogenesis factors correlated with the 40S and 60S profiles in Figure 4.4B and C respectively. The ribosomal proteins correlated with the profile in Figure 4.4D. Color indicates significance of chi-squared test, from black (most significantly different) to red (least significantly different).

4.3.2 Sucrose density gradient and mass spectrometry of Hela cell lysate

Hela cells were lysed and the cytoplasmic and nuclear fractions were loaded onto two sucrose density gradients. Mass spectrometry shotgun proteomics was applied to quantify the amount of each protein detected along the two sucrose density gradients. In total, 3,013 proteins were detected and quantified across 14 fractions from separation of cytoplasmic proteins and 14 fractions from separation of nuclear proteins. A clustergram of all the identified proteins was shown in Figure 4.6A. Several complexes were identifiable from the clustering analysis, such as ribosome, proteasome, TCP1 chaperonin complex, and NADH dehydrogenase cytochrome c oxidase complex (Figure 4.6A). The sedimentation profiles of the components in TCP1 chaperonin complex, RNA polymerase II complex, and NADH dehydrogenase 1b complex were shown in Figure 4.6B. The protein GRIM-19, initially identified as a regulator of cell death induced by interferon-beta and retinoic acid, was later identified as a subunit of the NADH dehydrogenase complex 1 (Fearnley et al., 2001). In the co-sedimentation analysis that was based on mass spectrometry, GRIM-19 exhibited similar co-sedimentation pattern as the other components in NADH dehydrogenase complex 1b, strongly supporting its association with other components in this complex (Figure 4.6B).

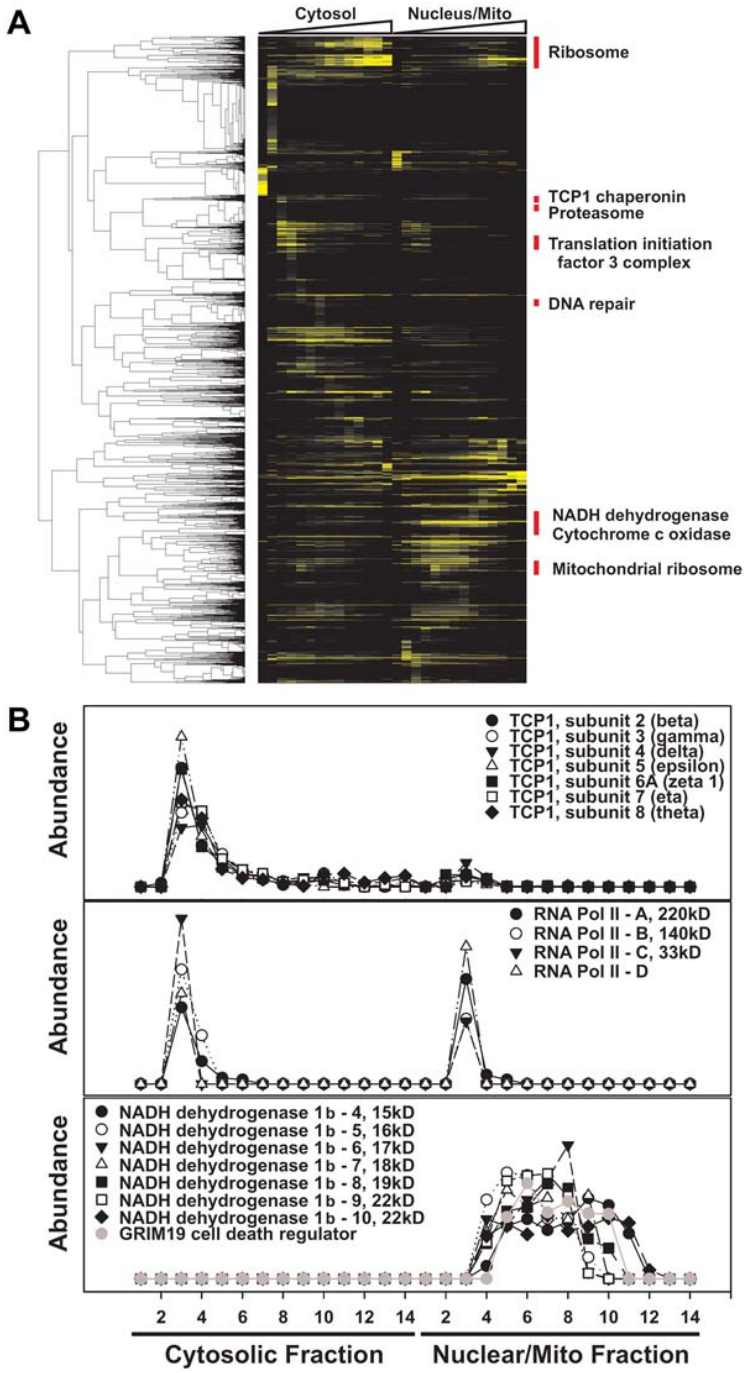


Figure 4.6 Protein sedimentation profiles for Hela cells measured by mass spectrometry.
(A) A clustergram of identified proteins by mass spectrometry from sucrose density gradient separation of cytoplasmic and nuclear fractions. Each row represents a protein and each column represents a fraction from the sucrose density gradient.
(B) Components of several known protein complexes exhibited similar sedimentation profiles.

4.4 DISCUSSION

By combining sucrose density gradient separation and mass spectrometry identification and quantification, protein components within known protein complexes were observed with similar sedimentation profiles along the gradient, suggesting that this approach could potentially be applied to identify protein complexes in a complex sample such as whole cell lysate. In this preliminary study, the sucrose density gradient (7-47%) was designed to separate large complexes such as proteasomes and ribosomes. A different gradient such as 5-20% sucrose gradient or 10-30% glycerol gradient could be used to separate small complexes.

The unique feature of this approach is to analyze protein complexes at their native states and its potential broad application to any organism. Additionally, the sizes of different complexes could be estimated based on the standards with known sizes. However, several limitations of this approach do exist. Some limitations might be reduced or eliminated by using other techniques.

First, due to the small number of fractions collected (14 fractions for each gradient in this study), some complexes with similar sizes and shapes might show similar sedimentation profiles and could not be distinguished from each other from the sedimentation profiles. Increasing the fractions collected might alleviate this problem. But there are a limited number of fractions that could be collected because of the limited separation power of the gradient. Another way to solve this problem is to separate the protein complexes by using other separation techniques based on different physical properties of the complexes such as the charge. Protein complexes with similar sizes and shapes might have different charges so that they could be separated out by ion-exchange columns or isoelectric focusing. Therefore, by using two separation techniques based on

different physical properties of the protein complexes, it is possible to identify many complexes from the cell lysate.

Second, a fraction of proteins in the cell lysate were identified by mass spectrometry. For example, 1,023 and 3,013 unique proteins were identified from the fractions collected from separation of yeast cell lysate and Hela cell lysate respectively, whereas over 4,000 yeast proteins could be detected by Western blot (Ghaemmaghami et al., 2003). This is partly because the samples used for separation are soluble fractions of the total cell lysate, representing a subset of the proteome. Another reason is because of the limitation of the mass spectrometer for complex sample analysis due to ion suppression by highly abundant proteins in the sample. Recent advances in instrumentation of mass spectrometer and extensive fractionation of digested peptides could possibly alleviate this problem. For example, linear ion trap combined with Orbitrap could greatly speed up the analysis of peptides and simultaneously increase the accuracy of the analysis so that a deep sampling of peptides could be attained to identify some low abundant peptides in the digested sample. Two dimensional separations of peptides could reduce the complexity of the peptide mixture such that low abundant peptides could be analyzed by mass spectrometer. Nonetheless, even a fraction of the proteome could be revealed, this approach to identifying protein complexes would be still useful in general, especially for human cells.

Chapter 5: Network-Guided Identification of Ribosome Biogenesis Genes

5.1 BACKGROUND

In eukaryotic cells, the synthesis of ribosomes is a complex process involving several hundred genes spanning transcription of ribosomal precursor RNAs (pre-rRNAs), translation and nuclear import of ribosomal proteins (r-proteins), processing of pre-rRNAs, assembly of r-proteins with pre-rRNAs, and nuclear export of the ribosomal particles (Fatica and Tollervey, 2002; Fromont-Racine et al., 2003; Granneman and Baserga, 2004; Tschochner and Hurt, 2003; Venema and Tollervey, 1999; Zemp and Kutay, 2007). This pathway has been extensively studied over the past 30-40 years and many details are well documented for the yeast *Saccharomyces cerevisiae*, producing a broad picture of the major events. First, the 35S polycistronic pre-rRNA is transcribed from the rDNA repeat by RNA polymerase I in the nucleolus. During transcription, r-proteins and other trans-acting factors are assembled with the 35S pre-rRNA to form a 90S particle. Subsequently, the 35S pre-rRNA is modified at multiple sites and cleaved to form the pre-60S and pre-40S particles that are further processed within both nucleolus and nucleoplasm and exported into cytoplasm through nuclear pore complex (NPC) for further maturation. Eventually the pre-rRNA is processed into three mature rRNAs: 18S rRNA to form the 40S small subunit with 32 additional r-proteins; 25S and 5.8S rRNAs to form the 60S large ribosomal subunit with 46 r-proteins and an additional 5S rRNA whose precursor is separately transcribed by RNA polymerase III.

Ribosome biogenesis is a temporally and spatially dynamic process requiring coordination of many trans-acting factors at different stages along this pathway, including about 76 different species of small nucleolus RNAs (snoRNAs), most of which facilitate post-transcriptional modifications of pre-rRNAs, and at least 150 protein factors acting to

modify and cleave pre-rRNAs and helping to assemble and export ribosomal particles (Kressler et al., 1999; Piekna-Przybylska et al., 2007; Zemp and Kutay, 2007). Many of the protein factors were first identified by yeast genetic screens. Later biochemical purifications not only identified more protein factors in the pre-ribosomal particles but also shed light on intermediate steps in ribosome biogenesis pathway (Bassler et al., 2001; Dragon et al., 2002; Grandi et al., 2002; Harnpicharnchai et al., 2001; Nissan et al., 2002; Schafer et al., 2003). In a large-scale effort to identify genes for noncoding RNA processing, 115 mutants were shown using oligonucleotide microarrays to exhibit pre-rRNA processing defects, and 10 new genes were confirmed to affect pre-rRNA processing (Peng et al., 2003). Despite these intensive studies, new ribosome biogenesis genes are still emerging and computational analysis suggests that more than 200 genes constitute the ribosome biogenesis regulon (Wade et al., 2006), indicating that genes in this fundamental cellular pathway have not been completely identified.

We asked if recent functional genomic and proteomic studies could be applied in a prediction fashion to identify candidate ribosomal biogenesis genes. In particular, functional networks of genes have been reconstructed, incorporating literally millions of experimental observations (e.g. DNA microarray data, protein interactions, comparative genomics, and literature mining) into probabilistic networks of genes, indicating genes likely to work together in cells. We used such a probabilistic gene network (Lee et al., 2007) to predict the yeast genes most likely to participate in ribosome biogenesis based upon connectivity to known ribosomal biogenesis genes. We present here experimental evaluation of 212 predictions, demonstrating involvement of about 30 new genes in the ribosomal biogenesis pathway.

5.2 MATERIALS AND METHODS

5.2.1 Strains

Haploid MATa deletion mutants (Giaever et al., 2002) were obtained from Research Genetics. $TetO_7$-promoter mutants (Mnaimneh et al., 2004) and TAP-tagged strains (Ghaemmaghami et al., 2003) were acquired from Open Biosystems. Mutants with GAL1-promoter controlled essential alleles were constructed in strain BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) by inserting the kan^r gene and the GAL1 promoter before the start codon of each gene via homologous recombination (Longtine et al., 1998). Mutant with GAL1 promoter controlled *PAP2* was constructed in strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) by inserting the His3 gene and the GAL1 promoter before the start codon of *PAP2* via homologous recombination. GAL1-promoter controlled mutants were confirmed by PCR. The oligos used to construct the GAL1-promoter controlled mutants were listed in Table 5.1.

To construct a mutant with deletion of *TRF5* and GAL1 promoter controlled *PAP2*, MATa *trf5Δ::kanMX6* mutant was mated with MATα *His3MX6:GAL1-PAP2* mutant to obtain a heterozygote, followed by tetrad dissection and selection to obtain the double mutant.

Table 5.1 Oligos for construction of GAL1 promoter controlled mutants

Name	**Sequence**
YIL091C-F4	TGCATTGGTTCAGTGAGACGGTGGTATAATTGCGAAGTGACCAAATTGTAgaattcgagctcgtttaaac
YIL091C-R2	TTCCGATAGCCACGGAAATTATCATTCTTTTCCCTCACAGAACTGTCACTcatttgagatccgggtttt
YIL091C-C	CATCGCCACTGTCGACATCG
YMR185W-F4	CTCATGTTTCTTCTTGTCATATCAGATCAAGAAAAACAAAAGGTTGGAAAgaattcgagctcgtttaaac
YMR185W-R2	CTTGTATTCAATATATCGTGTATGTTGATCTTTTGCTCTTTATCCTCATTcattttgagatccgggtttt
YMR185W-C	TATTTGAGCACTTCAACGCC
NOP9-F4	TACTGTGAACGACGCTGAAACCTTCACGGAAAAACCACATTATTGTTATTgaattcgagctcgtttaaac
NOP9-R2	TCCTTTCTCTGTTTGTCTTGGTGTCTTCTGCCTCTTGTTTTAGTCTTTCCcattttgagatccgggtttt
NOP9-C	CGGCTTGCTTGAAATACTCC
YRB2-F4	ATGTCAACAAATCAATTTGCACAAAACAAGTACGCATAAACTTTGAGAGCgaattcgagctcgtttaaac
YRB2-R2	TGTTTCACCTCAGAATTTTCCCTGGCTGCATTGCCACCATTGGTCTCACTcattttgagatccgggtttt
YRB2-C	TCACTCGTTTCTTCCGCTTG
SUA5-F4	GTTTCCATTTTTAATGGTAGTCCTGTTTATTATTCCATTTTTAGAATTGTgaattcgagctcgtttaaac
SUA5-R2	GTATCAAACAGTGCTTTCGATGTCATTGCCAAAAAATGTCGTCCAAGGTAcatttgagatccgggtttt
SUA5-C	5'- TTCAACTAGCGCTGCTTCCG -3'
SCC4-F4	GGGAATCAGTATTGGTAGAAATTGAACGATTTCGAAAGAAGTACAAATTCgaattcgagctcgtttaaac
SCC4-R2	TGTGCTAAGTGGTATACTTGTGAAATACTAAGTTTGTCGCCAAGATTTTCcatttgagatccgggtttt
SCC4-C	CGTGACCTTACTATCCTCCAG
ENP2-F4	CATTTATTACATCAGTTACAGGTATCAACATATATCTTCAAAAAGGGGACATTgaattcgagctcgtttaaac
ENP2-R2	GTACCAGAAACCTGATATACTGAAACATCATTTGCGGAAGTAGATTTCAAAACcattttgagatccgggtttt
ENP2-C	TGTATGTACCGGTAGCCATGC
YAE1-F4	TTTTGATAGAAGAGAACAAAAAAGGAACAGTGTTACAATAGTATATAGAGgaattcgagctcgtttaaac
YAE1-R2	TCAGTTTCGACATCACTGTCTGATGCCCATACATCGTCCCAAGTATTCGAcatttgagatccgggtttt
YAE1-C	CCATCATTGAAACCCTCTTGC
PAP2-F4	TTCATAAAGTGTGAATAAGCAAGGGAACTATACTTGAAATgaattcgagctcgtttaaac
PAP2-R2	TAATCTTCTTTGAAGAAGAGGCTGTTACACTCTTTGCCCCcattttgagatccgggtttt
PAP2-C	CTCTTGCTCATTACGACTAGGT

Note: -F4 is the forward primer for GAL1 promoter tagging; -R2 is the reverse primer for GAL1 promoter tagging; -C is the reverse primer for GAL1 promoter tagging confirmation; KanB (CTGCAGCGAGGAGCCGTAAT) is the forward primer for confirmation.

5.2.2 Polysome profile analyses

Haploid deletion mutants were cultured to OD_{600} 0.3-0.5 in YPD at the conditional temperature (20°C, 30°C, or 37°C). $TetO_7$-promoter mutants were cultured in YPD and then diluted into YPD with 10ug/ml doxycycline (Fisher Scientific) for 9-20 hrs to OD_{600} 0.3-0.5. GAL1-promoter mutants were cultured in YPGal and then diluted into YPD for 12-20 hrs to OD_{600} 0.3-0.5. 200 ug/ml cycloheximide (Sigma) was added to each culture. Cultures were immediately cooled with ice and all subsequent steps were performed on ice or at 4°C. Cells were collected by centrifugation at 3000g for 5 minutes and washed once with lysis buffer (20 mM Tris pH 7.4, 20 mM KCl, 5 mM $MgCl_2$, 200 ug/ml cycloheximide, 12 mM β-mercaptoethanol). The cell pellet was resuspended in the lysis buffer with protease inhibitors (2 ug/ml leupeptin, 2 ug/ml aprotinin, 1 ug/ml bestatin, 1 ug/ml pepstatin A, 1mM PMSF) and lysed by vortexing with glass beads. Crude lysates were clarified by centrifuging at 15,000g for 10 minutes. Fifteen OD_{260} units of the supernatant were loaded onto a continuous 7% to 47% sucrose gradient. After centrifuging for 2.5-hr at 40,000 rpm in a Beckman SW40 rotor, the sucrose gradient was fractionated and absorbance at 254 nm was measured (ISCO fractionator). Each mutant polysome profile was aligned to the wild type reference polysome profile using correlation optimized warping (COW) implemented in Matlab (Giorgio Tomasi, 2004). Aligned polysome profiles were hierarchically clustered using Cluster and Treeview (Eisen et al., 1998).

5.2.3 Immunoblot analysis

TAP-tagged strains were cultured in YPD at 30°C to OD_{600} 0.3-0.5 and subsequent steps were performed as for the polysome profile analyses. Fractions from the sucrose density gradient were collected and 25 ul of each fraction was deposited onto a

nitrocellulose membrane using a 96-well dot-blot system (Schleicher & Schuell). The membrane was probed for the TAP-tagged proteins with the rabbit peroxidase anti-peroxidase (PAP) soluble complex (Rockland Immunochemicals), using Luminol (Santa Cruz Biotechnology) as the substrate for detection. The total intensity of each dot was quantified with Quantity One 1-D Analysis Software (Bio-Rad).

5.2.4 Northern blots

RNA was extracted by the hot acidic phenol method. The high and low molecular weight RNA species were separated by 1% agarose-formaldehyde gel (NorthernMax, Ambion) and 8% polyacrylamide-TBE-urea gel, respectively. RNAs were transferred onto Zeta-Probe GT membrane (Bio-Rad) by capillary transfer for agarose gel or semi-dry electroblotting for polyacrylamide gel. After UV cross-linking of the RNAs to the membrane, 5'-P^{32}-labeled oligonucleotide probes (Table 5.2) were sequentially hybridized. Hybridization signals were detected by phosphorimaging and quantified using Quantity One (Bio-Rad). The logarithm ratio of total intensity of each RNA species from a mutant to that from the corresponding wild type was calculated and used for hierarchical clustering.

Table 5.2 Oligos for Northern blots

Oligo number	name	target	sequence
1	AJO249	5S	TCTGGTAGATATGGCCGCAACC
2	AJO190	18S	GTCTGGACCTGGTGAGTTTCCC
3	AJO130	20S	TCTTGCCCAGTAAAAGCTCTCATGC
4	AJO603	23S	TGTTACCTCTGGGCCCCGATTG
5	AJO313	35S, 27SA	TCCAGTTACGAAAATTCTTGTTTTTGACAA
6	AJO191	5.8S	CGCTGCGTTCTTCATCGATGCG
7	AJO282	7S	GGCCAGCAATTTCAAGTTA
8	AJO214	27S	GTTCGCCTAGACGCTCTCTTC
9	AJO192	25S	CCCGCCGTTTACCCGCGCTTGG
10	AJO317	scR1	CGTGTCTAGCCGCGAGGAAGGATTTGTTCC
11	AJO962	U2	GCGACCAAAGTAAAAGTCAAGAACGACTCCACAAG TGCGAGGGTCGCGAC

5.2.5 Ribosome export assay

Wild type strains or mutants were transformed with either pAJ907 (*RPL25-GFP CEN LEU2*) or pAJ1486 (*RPS2-GFP CEN LEU2*). Each strain was also transformed with pRS411-SIK1-mRFP (*SIK1-mRFP CEN MET15*). Strains were cultured in synthetic complete media minus leucine and methionine supplemented with 2% dextrose or 2% galactose. Essential gene expression was inactivated as for polysome profile analyses. Cells were fixed with 4% formaldehyde (Pierce) for 30 minutes and washed twice with PBS (pH 7.2). DAPI (Vector Laboratories) was used to stain DNA. Fluorescence was visualized using a Nikon E800 microscope and Photometrics CoolSNAP ES CCD camera.

5.3 RESULTS

5.3.1 Network-guided prediction of new ribosome biogenesis genes

A functional network of genes (Lee et al., 2007) that covers about 95% of yeast proteome was used to predict genes involved in ribosome biogenesis (Figure 5.1A). This network employs a probabilistic scoring scheme to quantitatively integrate currently available heterogeneous functional genomic and proteomic data sets including gene expression data across different conditions, protein-protein interaction data sets derived from literature curation, high-throughput yeast two-hybrid assay and affinity purification coupled with mass spectrometry, genetic interaction data, and *in silico* interaction data sets (Lee et al., 2007). Ribosome biogenesis genes were highly connected and predictable in this gene network, as shown by a plot of cross-validated true positive versus false positive prediction rate (Figure 5.1B). To predict new ribosomal biogenesis genes, we utilized gene connectivity information in the gene network, *i.e.* "guilt-by-association"

(Marcotte et al., 1999b; Walker et al., 1999) to known ribosome biogenesis genes, ranking each candidate gene by the naïve Bayesian probability of operating in the same pathway as the known ribosome biogenesis genes. From the top-scoring genes, 212 candidates were selected for experimental validation.

5.3.2 Conditional growth phenotypic analysis for non-essential genes

The synthesis of ribosomes is essential for cell growth and survival. In our candidate ribosome biogenesis gene list, 50 genes are essential and 162 genes are non-essential under standard laboratory culture conditions (Winzeler et al., 1999). If a non-essential gene affects ribosome biogenesis, deletion of the gene might be expected to reduce cell growth. We performed growth assays for each strain with a deletion of one of the 162 non-essential genes under three temperature conditions, 20°C, 30°C and 37°C (Figure 5.2). 52 mutants with conditional slow growth phenotypes and 50 mutants carrying conditional essential alleles were investigated further (Figure 5.1A)

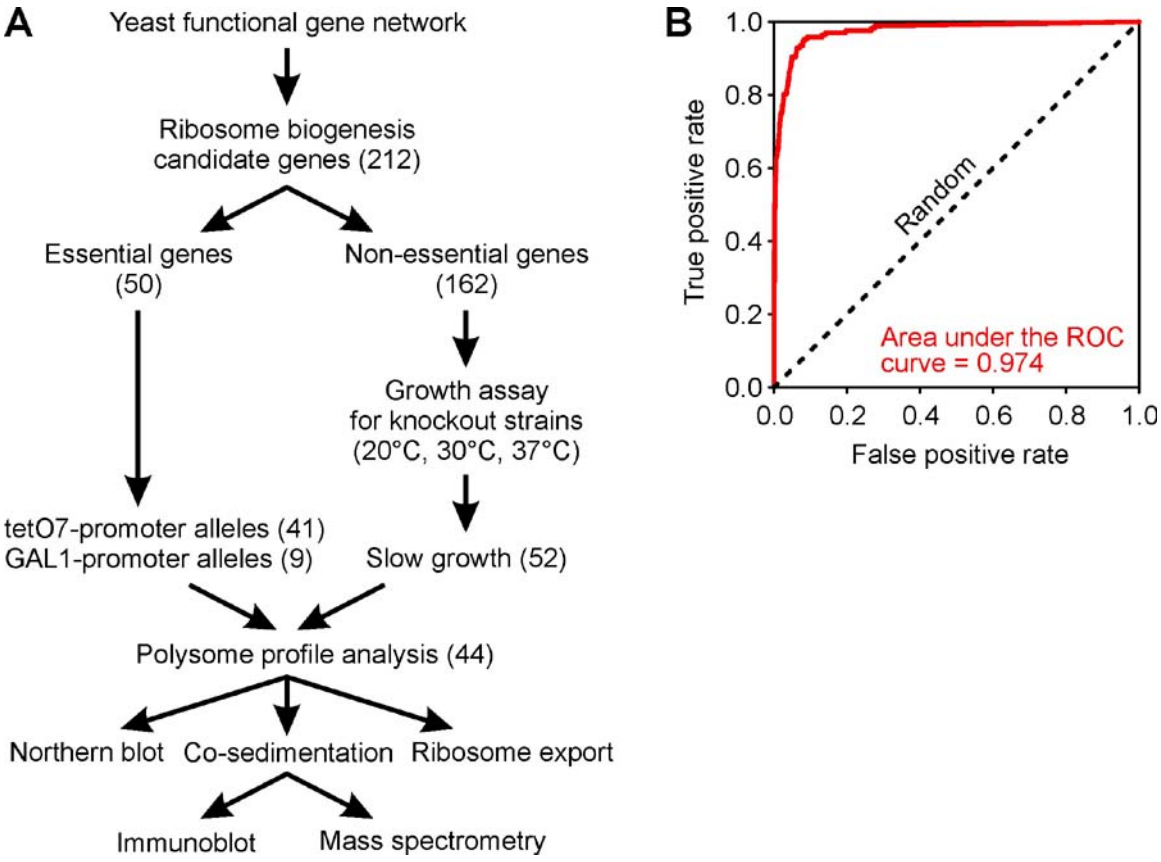
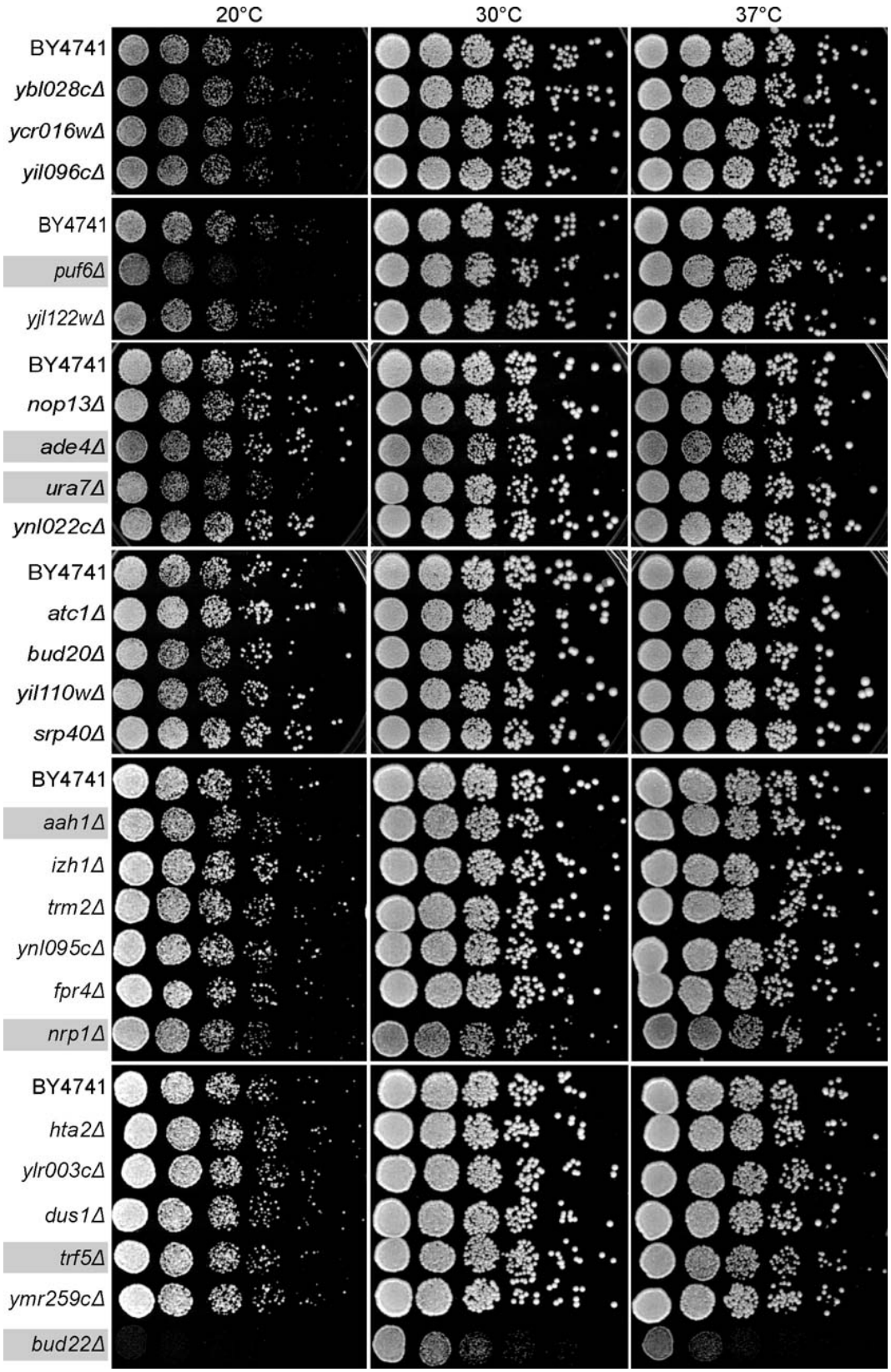
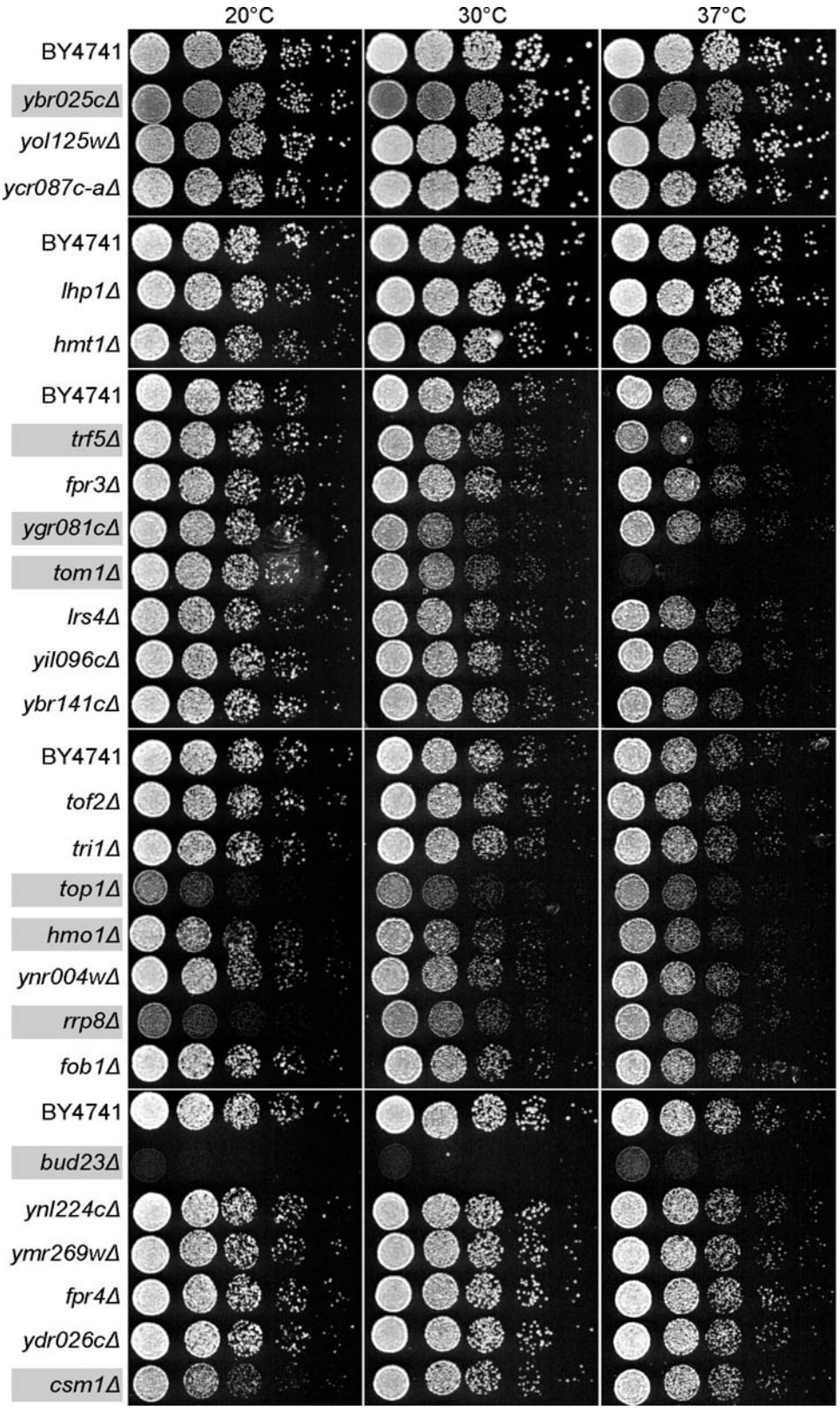
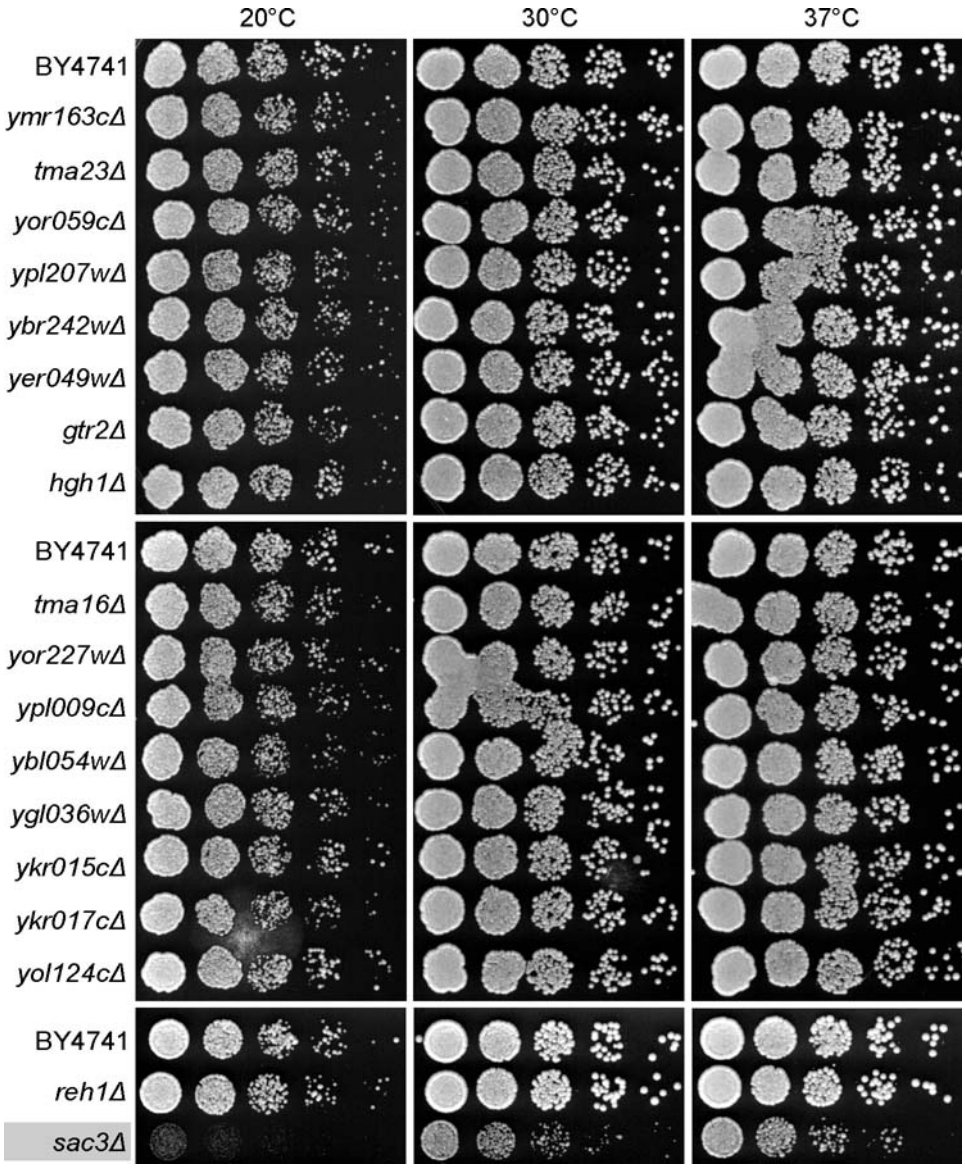


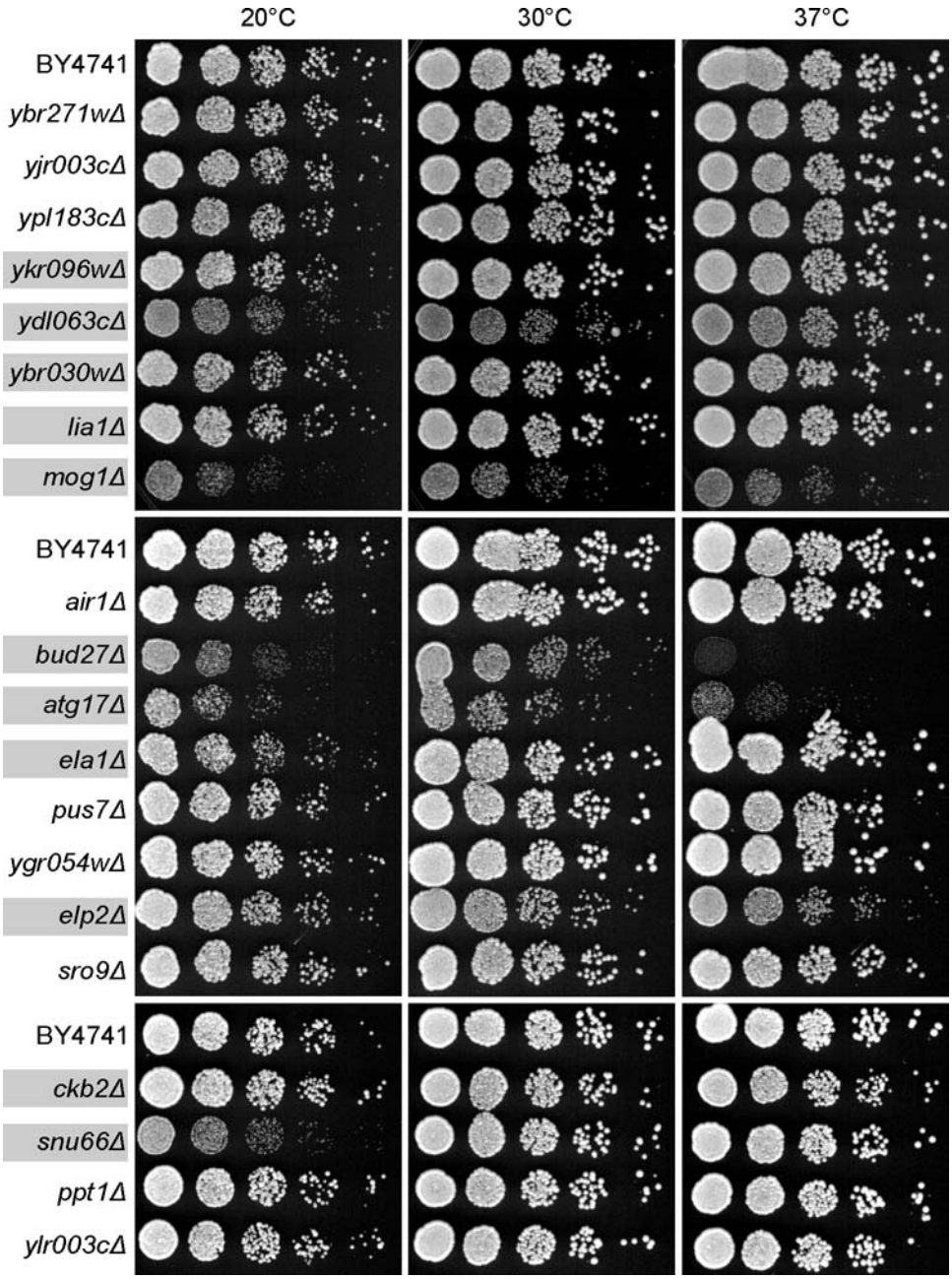
Figure 5.1. Overview of the analysis.
(A) A yeast functional gene network (Lee et al., 2007) reconstructed from diverse functional genomic and proteomic data was employed to predict genes for ribosome biogenesis. For non-essential genes, growth assays of the deletion mutants at different temperature conditions (20°C, 30°C, and 37°C) were used to identify conditional growth defects; polysome profiles of these strains were collected under slow growth conditions. For essential genes, mutants with conditional alleles were subjected to polysome profile analyses after depleting the encoded proteins. Genes affecting the ratio of free 40S to free 60S ribosomal subunits upon deletion of the gene or depletion of the encoded protein were further analyzed by co-sedimentation analyses to assign possible protein association with pre-ribosomal particles, by Northern blots to assay pre-rRNA processing defects, and by ribosomal subunit export assays.
(B) Assessment of the network-based predictability of ribosome biogenesis genes. The receiver-operator characteristic (ROC) curve (red line) indicated cross-validated recovery of known ribosome biogenesis genes based on their network connectivity to each other. True positive ribosome biogenesis genes were manually curated based on Gene Ontology annotation. The network-based prediction is considerably stronger than random expectation (dashed line).

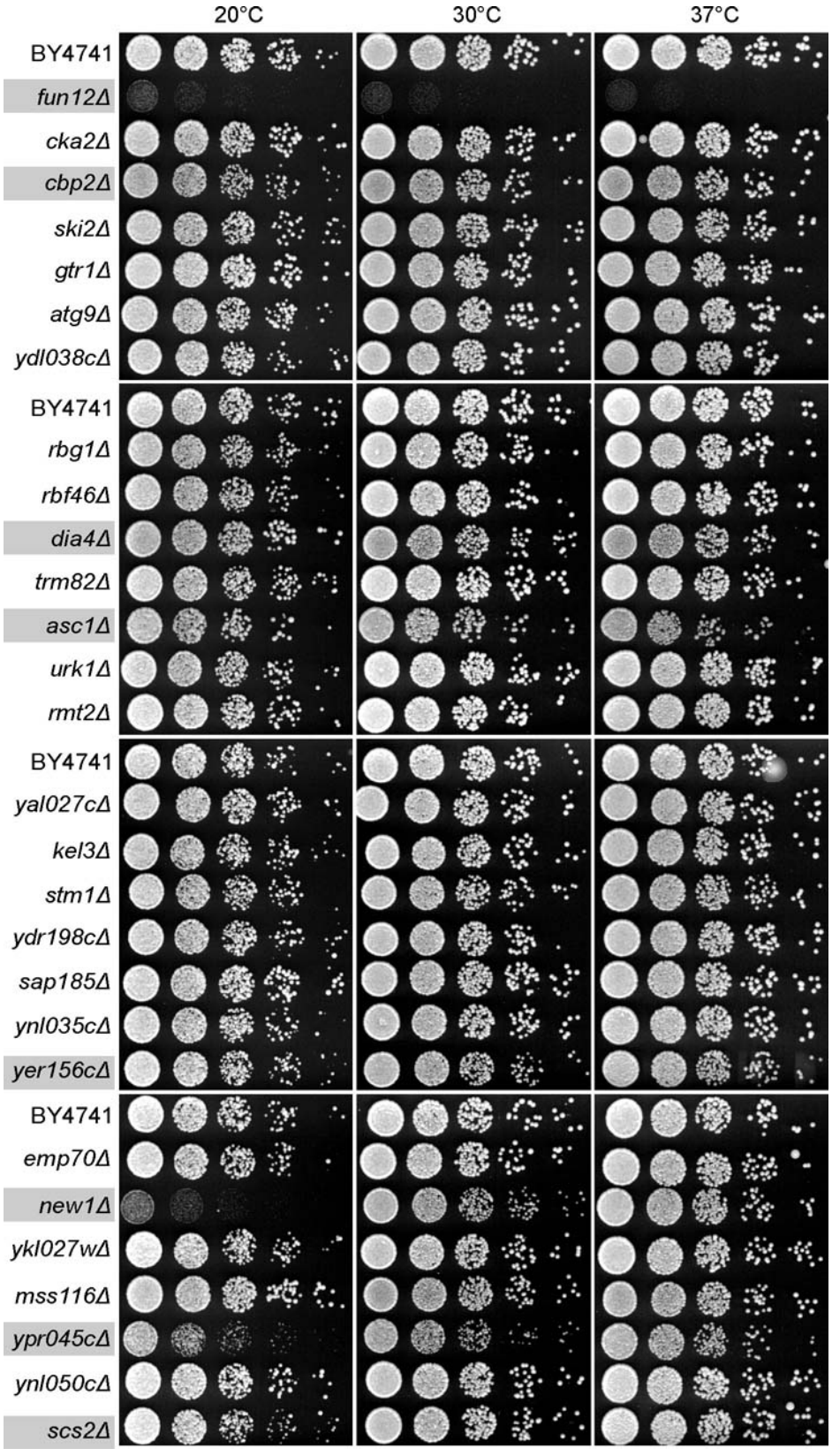
Figure 5.2 Growth assay for non-essential gene deletion mutants.
This figure is in the next seven pages. Deletion mutants were cultured in YPD and diluted to OD_{600} 0.1. A five-fold series dilutions were made for each mutant and 5ul diluted sample was deposited onto YPD plate. Mutants were then cultured at three different temperature conditions (20°C, 30°C, and 37°C). The mutants with slow growth phenotypes at any one of the conditions were highlighted with gray color.

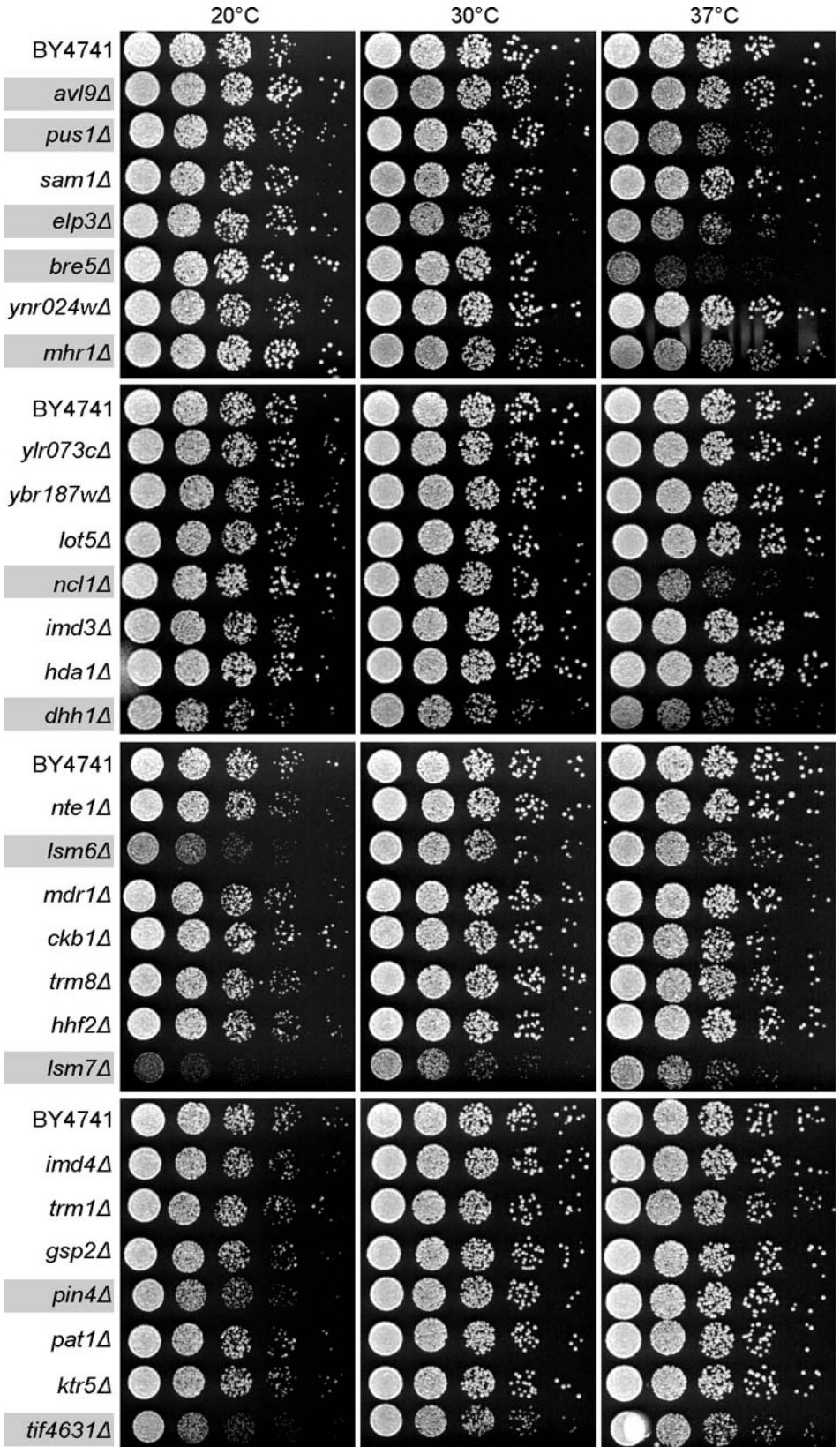

20°C
30°C
37°C
BY4741
ybl028cΔ
ycr016wΔ
yil096cΔ
BY4741
puf6Δ
yjl122wΔ
BY4741
nop13Δ
ade4Δ
ura7Δ
ynl022cΔ
BY4741
atc1Δ
bud20Δ
yil110wΔ
srp40Δ
BY4741
aah1Δ
izh1Δ
trm2Δ
ynl095cΔ
fpr4Δ
nrp1Δ
BY4741
hta2Δ
ylr003cΔ
dus1Δ
trf5Δ
ymr259cΔ
bud22Δ

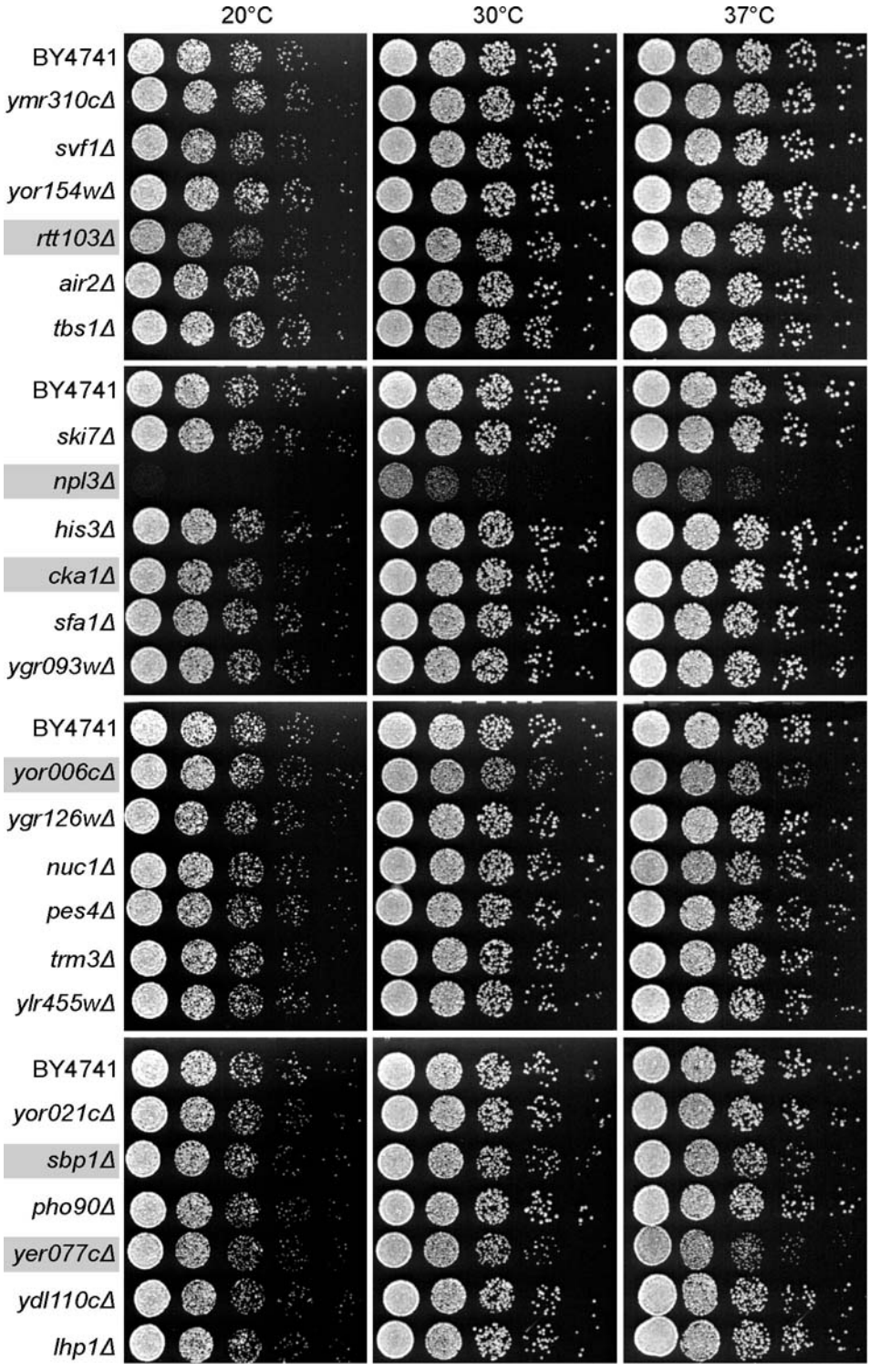

20°C
30°C
37°C
BY4741
ybr025cΔ
yol125wΔ
ycr087c-aΔ
BY4741
lhp1Δ
hmt1Δ
BY4741
trf5Δ
fpr3Δ
ygr081cΔ
tom1Δ
lrs4Δ
yil096cΔ
ybr141cΔ
BY4741
tof2Δ
tri1Δ
top1Δ
hmo1Δ
ynr004wΔ
rrp8Δ
fob1Δ
BY4741
bud23Δ
ynl224cΔ
ymr269wΔ
fpr4Δ
ydr026cΔ
csm1Δ


20°C
30°C
37°C
BY4741
ymr163cΔ
tma23Δ
yor059cΔ
ypl207wΔ
ybr242wΔ
yer049wΔ
gtr2Δ
hgh1Δ
BY4741
tma16Δ
yor227wΔ
ypl009cΔ
ybl054wΔ
ygl036wΔ
ykr015cΔ
ykr017cΔ
yol124cΔ
BY4741
reh1Δ
sac3Δ


20°C
30°C
37°C
BY4741
ybr271wΔ
yjr003cΔ
ypl183cΔ
ykr096wΔ
ydl063cΔ
ybr030wΔ
lia1Δ
mog1Δ
BY4741
air1Δ
bud27Δ
atg17Δ
ela1Δ
pus7Δ
ygr054wΔ
elp2Δ
sro9Δ
BY4741
ckb2Δ
snu66Δ
ppt1Δ
ylr003cΔ


20°C
30°C
37°C
BY4741
fun12Δ
cka2Δ
cbp2Δ
ski2Δ
gtr1Δ
atg9Δ
ydl038cΔ
BY4741
rbg1Δ
rbf46Δ
dia4Δ
trm82Δ
asc1Δ
urk1Δ
rmt2Δ
BY4741
yal027cΔ
kel3Δ
stm1Δ
ydr198cΔ
sap185Δ
ynl035cΔ
yer156cΔ
BY4741
emp70Δ
new1Δ
ykl027wΔ
mss116Δ
ypr045cΔ
ynl050cΔ
scs2Δ


20°C
30°C
37°C
BY4741
avl9Δ
pus1Δ
sam1Δ
elp3Δ
bre5Δ
ynr024wΔ
mhr1Δ
BY4741
ylr073cΔ
ybr187wΔ
lot5Δ
ncl1Δ
imd3Δ
hda1Δ
dhh1Δ
BY4741
nte1Δ
lsm6Δ
mdr1Δ
ckb1Δ
trm8Δ
hhf2Δ
lsm7Δ
BY4741
imd4Δ
trm1Δ
gsp2Δ
pin4Δ
pat1Δ
ktr5Δ
tif4631Δ


20°C
30°C
37°C
BY4741
ymr310cΔ
svf1Δ
yor154wΔ
rtt103Δ
air2Δ
tbs1Δ
BY4741
ski7Δ
npl3Δ
his3Δ
cka1Δ
sfa1Δ
ygr093wΔ
BY4741
yor006cΔ
ygr126wΔ
nuc1Δ
pes4Δ
trm3Δ
ylr455wΔ
BY4741
yor021cΔ
sbp1Δ
pho90Δ
yer077cΔ
ydl110cΔ
lhp1Δ

5.3.3 Identifying ribosomal subunit biogenesis defects by polysome profile analysis

The synthesis of 40S and 60S subunits are largely independent (Granneman and Baserga, 2004). Depletion of the factors required for the synthesis of either subunit usually does not affect the synthesis of the other subunit (Dragon et al., 2002; Hong et al., 2001), resulting in a change in the ratio of free 40S to free 60S subunits in the cell. This ratio will be lower in cells with defective 40S subunit biogenesis and higher for cells defective for 60S subunit biogenesis. Additionally, a reduction in the amounts of 60S subunits can lead to 40S subunits being stalled on actively translating mRNAs, awaiting 60S subunits in order to form competent ribosomes for translation. These stalled 40S subunits are observable as halfmer polysomes in a polysome profile (Adams et al., 2002; Lee et al., 2002). A typical polysome profile is generated by separating the ribosomal subunits and different size polysomes through a continuous sucrose density gradient and monitoring the absorbance of nucleic acids along the sucrose gradient (Warner et al., 1963). We performed polysome profile analyses for the 50 mutants carrying essential alleles controlled by either a tetracycline-regulatable ($tetO_7$) promoter (Mnaimneh et al., 2004) or a GAL1 promoter and for the 52 non-essential gene deletion mutants with conditional growth defects. To observe the effects of essential genes on ribosome biogenesis, each mutant was cultured either by adding doxycycline into the medium or changing the carbon source of the medium from galactose to dextrose in order to shut down transcription of the essential gene. Mutants carrying null alleles were cultured under the conditions showing slow growth phenotypes. The ribosome subunits and the polysomes in the mutants were assayed as polysome profiles.

Including controls, over 150 polysome profiles were generated. In order to compare different profiles and perform multivariate analyses such as clustering, we computationally aligned each profile to a reference wild type profile by using a

correlation optimized warping algorithm (Giorgio Tomasi, 2004) which corrects for peak shifts of ribosome subunits and polysomes due to minor variations in sucrose density gradients. Similar polysome profiles were grouped together using hierarchical clustering (Eisen et al., 1998). From the clustergram, the signals corresponding to the ribosomal subunits, mono-ribosomes, polysomes, and halfmer polysomes were clearly identifiable (Figure 5.3). Most of the profiles with high 40S to 60S ratios and halfmer peaks were in clusters 1 and 2, representing 60S biogenesis defects (Figure 5.4B). Cluster 3 represented profiles from mutants showing protein translation defects (Figure 5.4C). The profiles with low 40S to 60S ratios were in cluster 4, suggesting 40S biogenesis defects (Figure 5.4D). The polysome profiles from three mutants, *ypr045cΔ*, *tif4631Δ*, and *snu66Δ*, were not clustered with 60S biogenesis cluster 1 and 2, although they showed halfmer polysomes and imbalance of 40S and 60S subunits compared to wild type strains (Figure 5.3, 5.4B). Some mutants showed only subtle defects and their profiles were interspersed among wild type-like profiles during clustering (Figure 5.5). The polysome profiles of mutants provided initial suggestions as to the general natures of their involvements in ribosome biogenesis and translation process. We further investigated 44 mutants that exhibited imbalanced 40S to 60S ratios compared to wild type strains (Figure 5.1A).

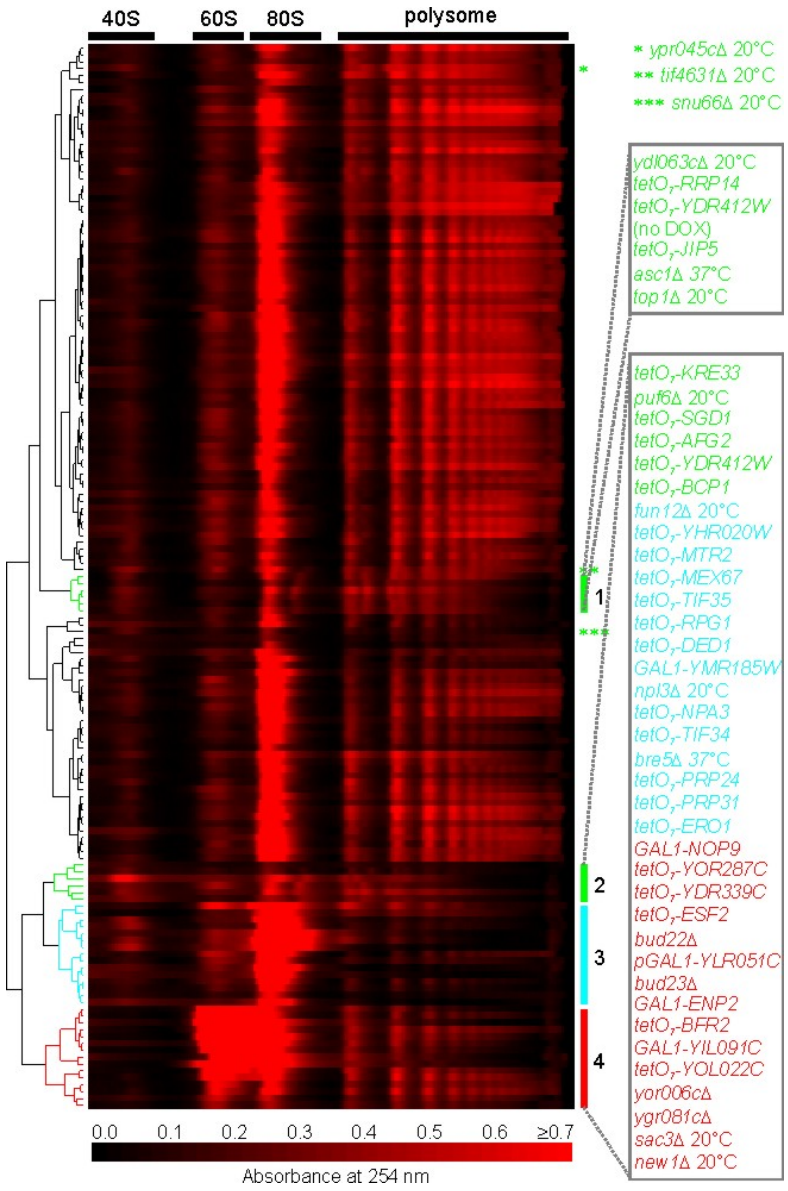


Figure 5.3 Hierarchical clustering of polysome profiles of the mutants.
Cluster 1 and 2 represent mutants with 60S subunit biogenesis defects (green labels), cluster 3 represents mutants with translation defects (cyan labels), and cluster 4 represents mutants with 40S subunit biogenesis defects. Three additional mutants with 60S subunit biogenesis defects are labeled with stars. Each row corresponds to the polysome profile of a single strain, indicating nucleic acid absorbance as a function of position along a sucrose density gradient. Strains were cultured at 30°C unless otherwise indicated. Mutants with $tetO_7$-promoter alleles were cultured in medium with 10ug/ml doxycycline (+DOX) unless indicated with no DOX. Mutants with GAL1-promoter alleles were first cultured in medium with galactose (Gal) as the carbon source and then diluted into medium with dextrose (Glc) as the carbon source.

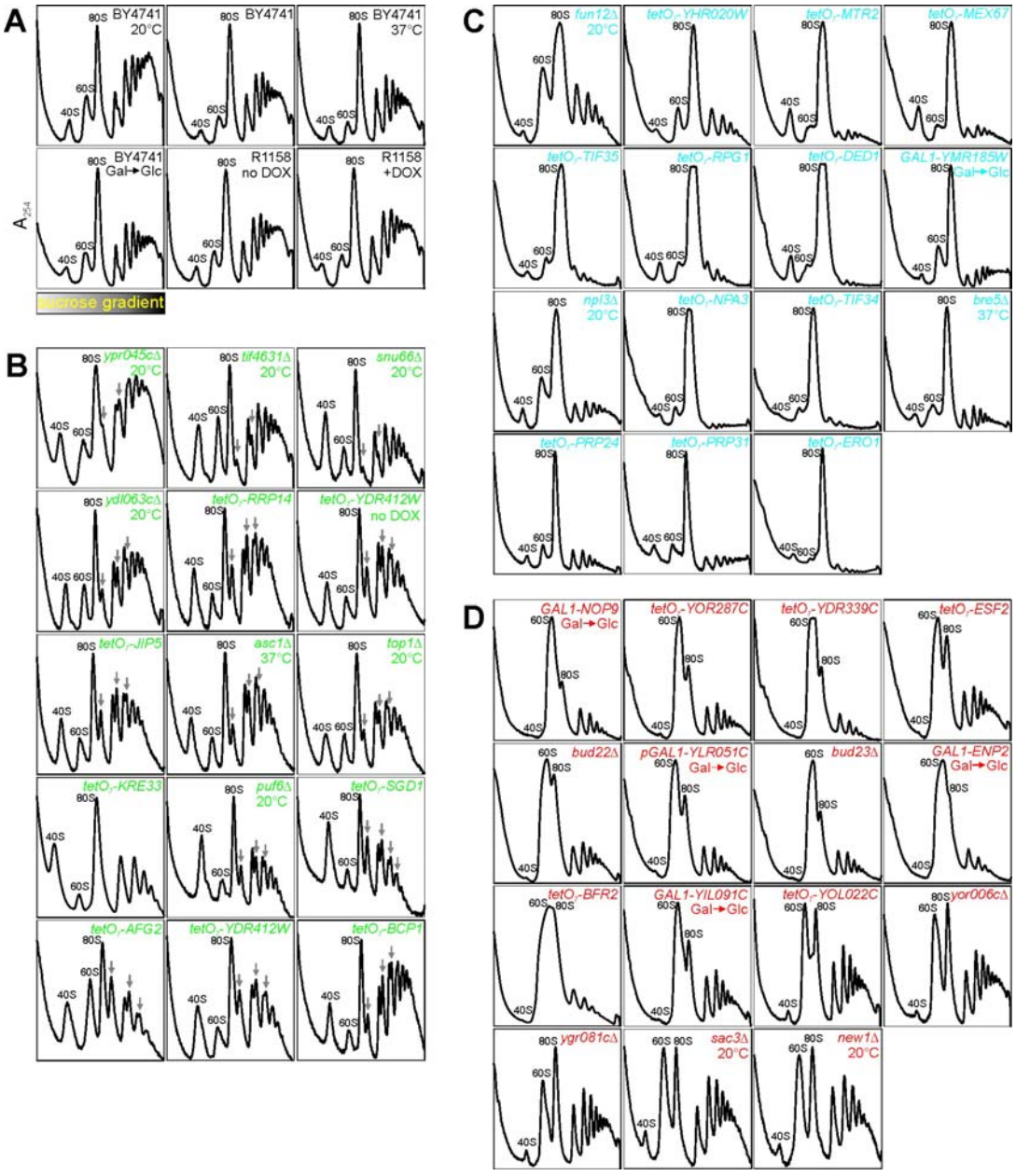


Figure 5.4 Polysome profiles of mutants shown in Figure 5.3.
(A) Polysome profiles of wild type strains cultured under assayed conditions. BY4741 is the control strain for the non-essential gene deletion mutants and mutants with GAL1-promoter alleles. R1158 is the control strain for the mutants with $tetO_7$-promoter alleles. Peaks corresponding to the 40S and 60S ribosomal subunits and 80S mono-ribosomes in the polysome profiles are labeled.
(B, C, and D) Polysome profiles of mutants with 60S subunit biogenesis defects, translation defects, and 40S subunit biogenesis defects. Gray arrows indicate halfmer polysome peaks.

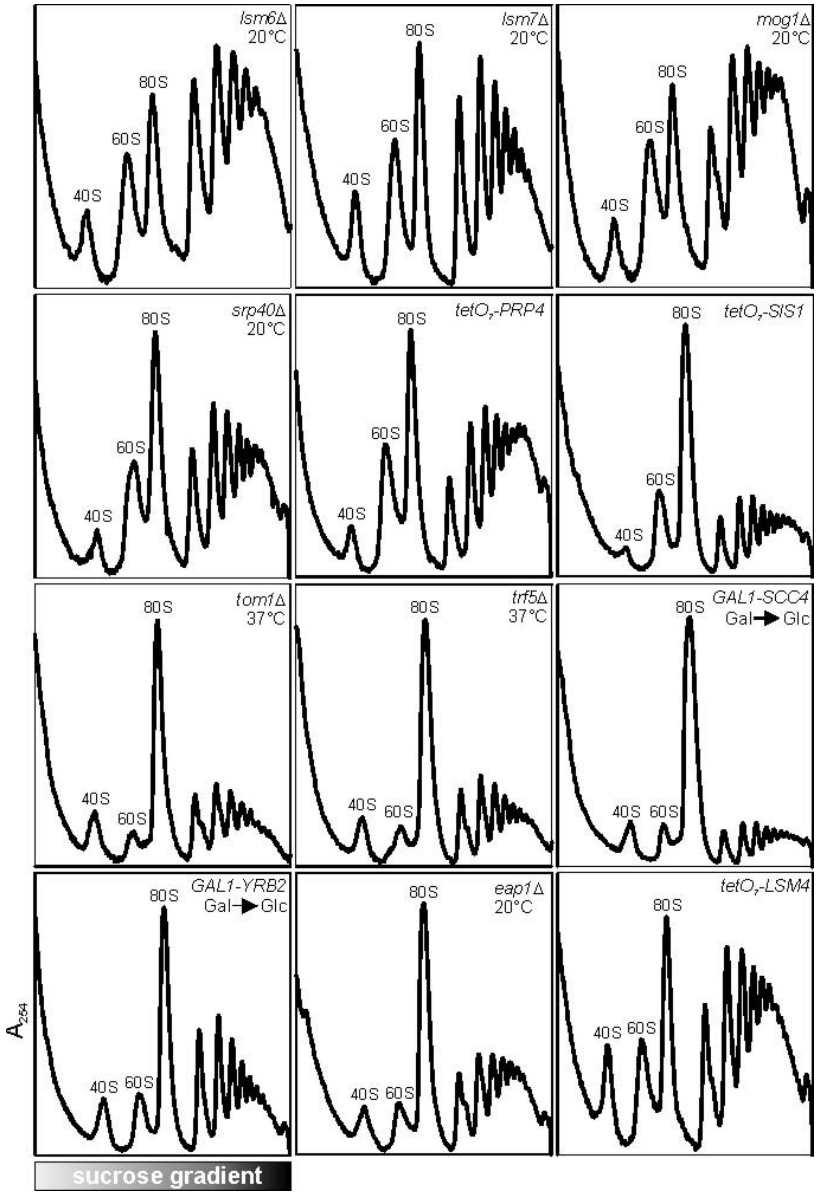


Figure 5.5 Polysome profiles of mutants with minor defects compared to control strains.

5.3.4 Co-sedimentation analysis with sucrose density gradient

Most ribosome biogenesis factors associate with pre-rRNAs and ribosomal proteins to form pre-ribosomal particles (Trapman et al., 1975; Tschochner and Hurt, 2003), some of which have been purified and partially characterized (Harnpicharnchai et al., 2001; Nissan et al., 2002; Schafer et al., 2003). To simply distinguish the factors associated with pre-40S from the factors associated with pre-60S, we applied a classical immunoblot approach to assessing sedimentation patterns of potential ribosome biogenesis factors along sucrose density gradient (Figure 5.1A).

Strains carrying TAP-tagged alleles (Tandem Affinity Purification tag) for 34 ribosome biogenesis candidates with polysome profile defects were available (Ghaemmaghami et al., 2003) and used to prepare samples for sucrose density gradients. After separating the ribosomal subunits and polysomes, fractions of each sucrose gradient were collected and analyzed for the TAP-tagged protein by immunoblot (Figure 5.6A). The abundance of each tagged protein within each fraction was quantified by the total intensity of each dot (Figure 5.6B-F) (Borggrefe et al., 2001). We expected 40S biogenesis factors would mainly distribute in the 40S fractions (e.g., Tsr1-TAP in Figure 5.6A) and/or 80S/90S fractions if the protein associates with the pre-40S and/or 90S particles. 60S biogenesis factors would mainly distribute in the 60S fractions (e.g., Lsg1-TAP in Figure 5.6A). The r-proteins would be expected to distribute not only in the 40S or 60S fractions, but also in the mono-ribosome and polysome fractions (e.g., Rps3-TAP and Rpl8a-TAP in Figure 5.6A). By contrast, Eno1p, a cytosolic metabolic enzyme, was distributed in the low density fractions of a sucrose gradient (Figure 5.6A). We did not detect background signals from the wild type control strain under these experimental conditions (BY4741 in Figure 5.6A).

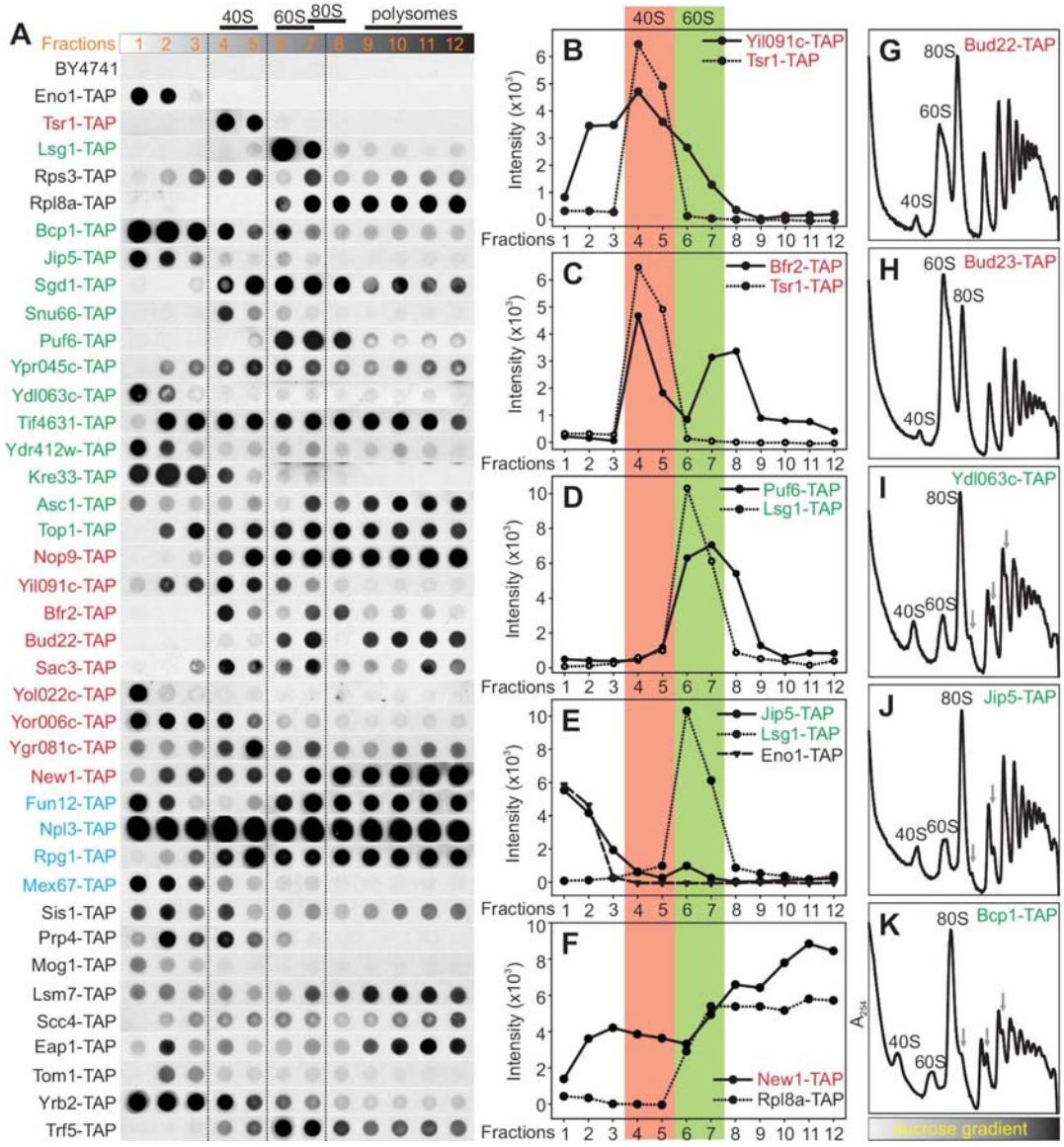


Figure 5.6 Co-sedimentation assay by sucrose density gradient and immunoblot.
(A) Immunoblots of fractions collected from sucrose density gradients for strains carrying TAP-tagged alleles of the labeled genes. Fractions 4 and 5, 6 and 7, 7 and 8, and 9-12 correspond to the 40S, 60S, 80S, and polysome peaks in the sucrose density gradients. BY4741 is the negative control. Tsr1-TAP and Lsg1-TAP are the positive controls for 40S subunit biogenesis factors and 60S subunit biogenesis factors, respectively. Rps3-TAP and Rpl8a-TAP show the locations of small and large ribosomal subunits in the sucrose density gradient, respectively, whereas Eno1-TAP represents the proteins that do not co-sediment with ribosomes.
(B-F) show the quantitation of the immunoblots for Yil091c-TAP, Bfr2-TAP, Puf6-TAP, Jip5-TAP, and New1-TAP.
(G-K) show the polysome profile defects for several TAP-tagged strains.

We observed that many proteins sedimented in either 40S or 60S fractions. Yil091cp, an uncharacterized protein, was enriched in 40S fractions (Figure 5.6B). Bfr2p was enriched in 40S fractions and 80S/90S fractions suggesting that this protein exists in both 40S pre-ribosomes and 90S pre-ribosomes (Figure 5.6C). Puf6p sedimented in 60S fractions (Figure 5.6D) supporting the 60S biogenesis defects observed in the polysome profile of *puf6Δ* (Figure 5.4B). Nop9p, a nucleolus localized protein (Huh et al., 2003), was enriched not only in 40S fractions, but also across all high density fractions (Figure 5.6A). This phenomenon also occurred for several other nucleolus localized proteins including Bud22p, Sgd1p, and Top1p (Figure 5.6A). Not surprisingly, translation initiation factors such as Tif4631p, Fun12p, Rpg1p, and Eap1p were highly enriched in the polysome fractions (Figure 5.6A). Interestingly, New1p, a protein involved in prion formation (Osherovich and Weissman, 2001), was observed across all fractions, with enrichment in high density fractions (Figure 5.6F). Several proteins (Jip5p, Ydl063cp, Ydr412wp, Kre33p, Yol022cp, and Yor006cp) shown to cause clear ribosome biogenesis defects following deletion of the gene or depletion of the protein (Figure 5.4B, 5.4D) mainly distributed in the low density fractions (Figure 5.6A), with Jip5-TAP showing only weak enrichment in the 60S fractions (Figure 5.6E). Although these proteins might transiently interact with pre-ribosomes or affect ribosome biogenesis indirectly, another explanation is that the carboxyl terminal tag might partially disrupt interactions between the ribosome biogenesis factors and the pre-ribosomes. We tested this latter case by assaying for ribosome biogenesis defects in strains with the TAP-tagged alleles. In several cases, we did observe the tag to confer ribosome biogenesis defects (Ydl063c-TAP, Jip5-TAP, and Bcp1-TAP in Figure 5.6I-K), with Jip5-TAP showing only minor defects for 60S biogenesis (Figure 5.6J). Additionally, Bud22p and Bud23p with carboxyl-terminal TAP tag also caused ribosome biogenesis defects (Figure 5.6G-H).

These observations suggest that the carboxyl-terminal TAP tag on some of the proteins may affect their interactions with other proteins and their proper functions in the cell.

5.3.5 New genes affecting pre-rRNA processing

Most ribosome biogenesis factors participate in ribosome assembly and/or pre-rRNA processing (Kressler et al., 1999). Pre-rRNA processing consists of sequential events involving multiple endo- and exo-nuclease cleavages (Venema and Tollervey, 1999). The 35S pre-rRNA is first cleaved at sites A_0, A_1, and A_2 to yield 20S species and $27SA_2$ species (Figure 5.7B). The 20S pre-rRNA is further processed in the cytoplasm to form the mature 18S rRNA after cleavage at D position. The $27SA_2$ pre-rRNA is processed by two different routes. Most of the $27SA_2$ species are cleaved at site A_3 followed by exonuclease digestion on both sides to the sites B_{1S} and B_2 to form $27SB_S$, while a small amount of $27SA_2$ are cleaved at B_{1L} and digested by exonuclease to B_2 to generate $27SB_L$. Then both 27SB species are processed at sites C_1 and C_2 to yield the mature 25S species and 7S species followed by exonuclease digestion to E to form the mature 5.8S (Figure 5.7B). Mutation of the gene or depletion of the protein required for ribosome biogenesis often leads to defective pre-rRNA processing directly or indirectly resulting in the reduction and/or accumulation of different pre-rRNA species (Venema and Tollervey, 1999).

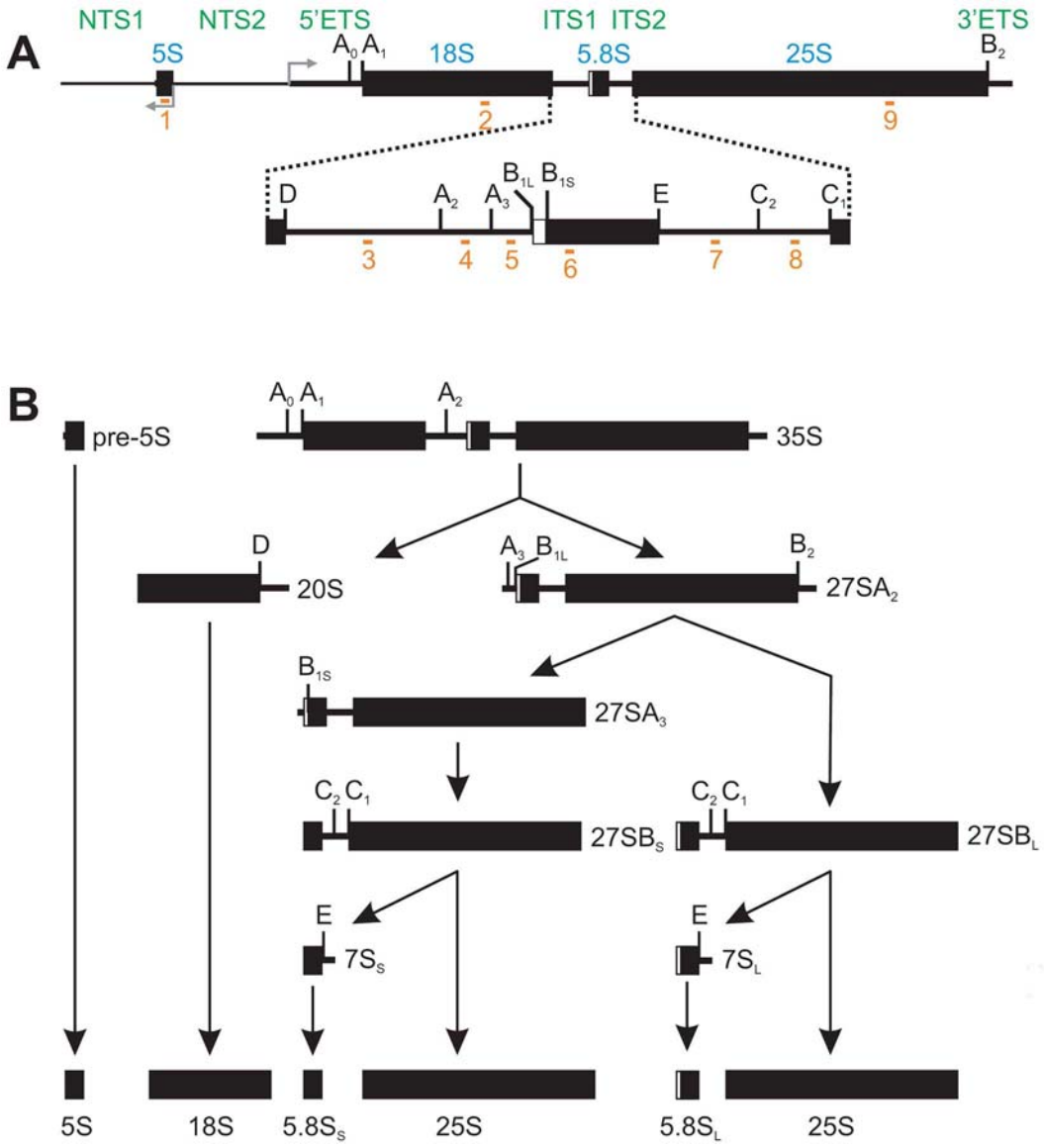


Figure 5.7 The rDNA repeat and pre-rRNA processing pathway.
(A) Oligo nucleotide probes (orange numbers) within an rDNA repeat were selected to probe the majority of pre-rRNA and rRNA species generated during the pre-rRNA processing pathway, diagrammed in (B).

To examine the detailed effects of the ribosome biogenesis candidate genes on pre-rRNA processing, we detected 9 different pre-rRNA and rRNA species by Northern blots in mutants either harboring a deletion of the gene or depleted for the protein. In order to quantitatively analyze the change of each RNA species in a mutant relative to its corresponding wild type, the intensity of each RNA species from the Northern blots was quantified and the logarithm of the intensity ratio of each RNA species from a mutant strain relative to that from its corresponding wild type strain was calculated. Figure 5.8 shows hierarchical clustering analysis of the Northern blots for the 44 mutants. We observed a dramatic increase (red signal in Figure 5.8) or decrease (green signal in Figure 5.8) of at least one pre-rRNA species for all of the mutants except *eap1Δ* and *trf5Δ* (Figure 5.8). Most mutants that showed 60S subunit biogenesis defects from the polysome profile analyses exhibited decreased levels of 25S rRNAs and increased levels of 35S pre-rRNAs. Many of those mutants also exhibited increased levels of 7S pre-rRNAs and clustered together, whereas *tetO7-BCP1*, *tetO7-JIP5*, and *tetO7-SGD1* clustered together showing little change of 7S, a minor increase of 7S, and a decrease of 7S respectively (Figure 5.8). Most 40S subunit biogenesis mutants also clustered together and exhibited increased levels of 35S pre-rRNAs and decreased levels of both 20S pre-rRNAs and 18S rRNAs (Figure 5.8).

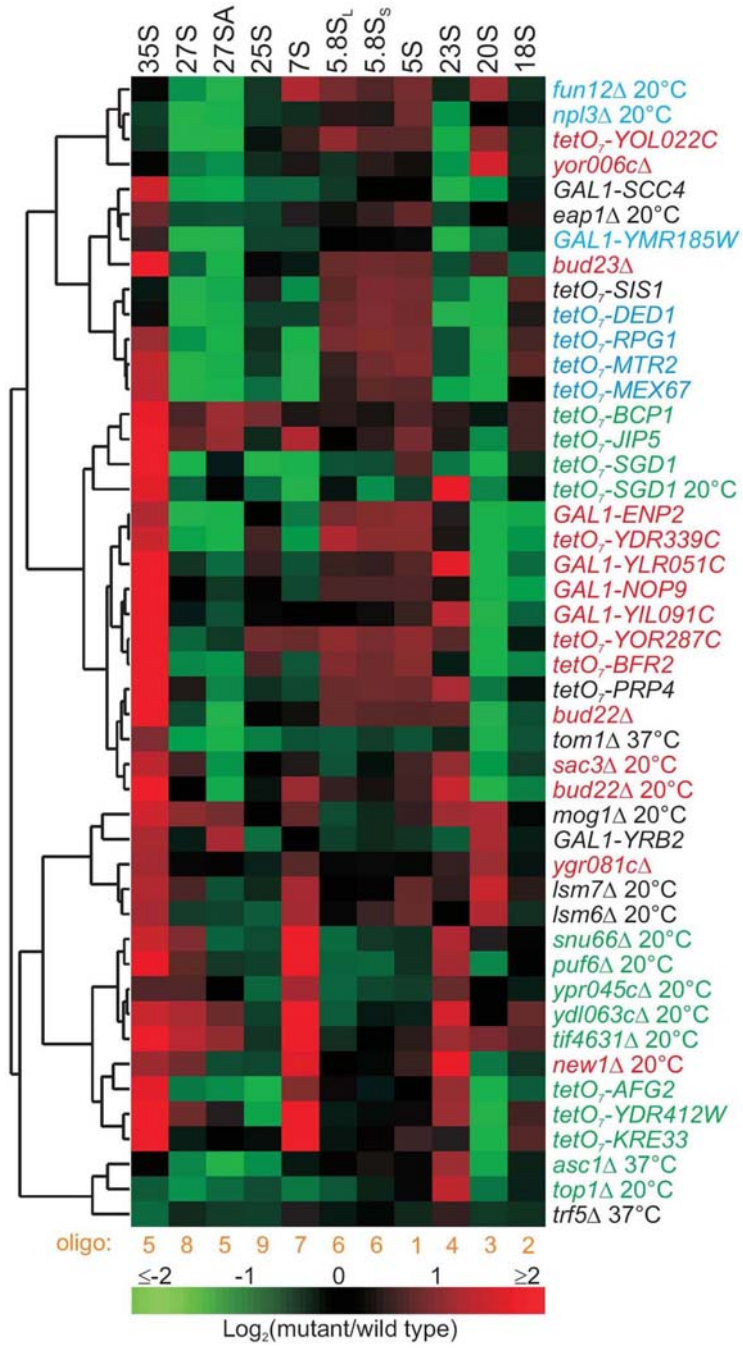


Figure 5.8 Hierarchical clustering of mutant strains on the basis of pre-rRNA abundances measured from Northern blots.

Each RNA species from Northern blots (Figure 5.9) was quantified, and the base 2 logarithm of mutant to wild type strain for each RNA species was calculated and used for clustering. Red and green colors represent increased and decreased levels of RNA species, respectively. Strain label colors are as in Figure 5.4.

5.3.5.1 Genes required for processing 35S pre-rRNA

The 35S pre-rRNA is modified by numerous snoRNPs and cleaved at the sites A_0, A_1, and A_2 to produce 20S and $27SA_2$, the pre-rRNAs for 40S small subunit and 60S large subunit, respectively (Figure 5.7B). Mutation or depletion of factors required for this processing step often leads to the accumulation of the 35S pre-rRNA and the reduction of the 20S and $27SA_2$ species (Venema and Tollervey, 1999). The lack of the 20S pre-rRNA leads to the reduction of the 18S mature rRNA resulting in reduced synthesis of the 40S small subunit. However, processing at site A_3 on the 35S pre-rRNA can still yield the $27SA_3$ species that is subsequently processed to the mature 5.8S and 25S for the 60S subunit (Gallagher et al., 2004). This type of 35S processing defect was observed upon mutation or depletion of Enp2p, Bfr2p, Ydr339cp, Ylr051cp, Yil091cp, Yor287cp, Nop9p, or Bud22p, all of which localize in the nucleolus where the 35S pre-rRNA is processed, except Yor287cp that lacks localization information (Huh et al., 2003). Ydr339cp (Fcf1p/Utp24p) and Ylr051c (Fcf2p) were recently confirmed as 40S biogenesis factors during the progress of this work (Rempola et al., 2006) and Ydr339cp was also observed to be a component of the SSU (small subunit) processome as a putative nuclease for the cleavage of the A_1 and/or A_2 sites, although direct evidence was lacking for the nuclease activity (Bleichert et al., 2006). Enp2p and Bfr2p did not co-immunoprecipitate with Mpp10p and U3 snoRNA, suggesting that they were not components of the SSU processome (Bernstein et al., 2004). However, Bfr2p sedimented with both 40S and 80S/90S fractions (Figure 5.6A), suggesting that it was a possible component of the large pre-ribosomal particles, presumably the earliest 90S complex. *YIL091C*, *YOR287C*, and *NOP9* are uncharacterized, and *BUD22* is known to affect bud-site selection (Ni and Snyder, 2001). We showed here that these four genes participate in 35S pre-rRNA processing.

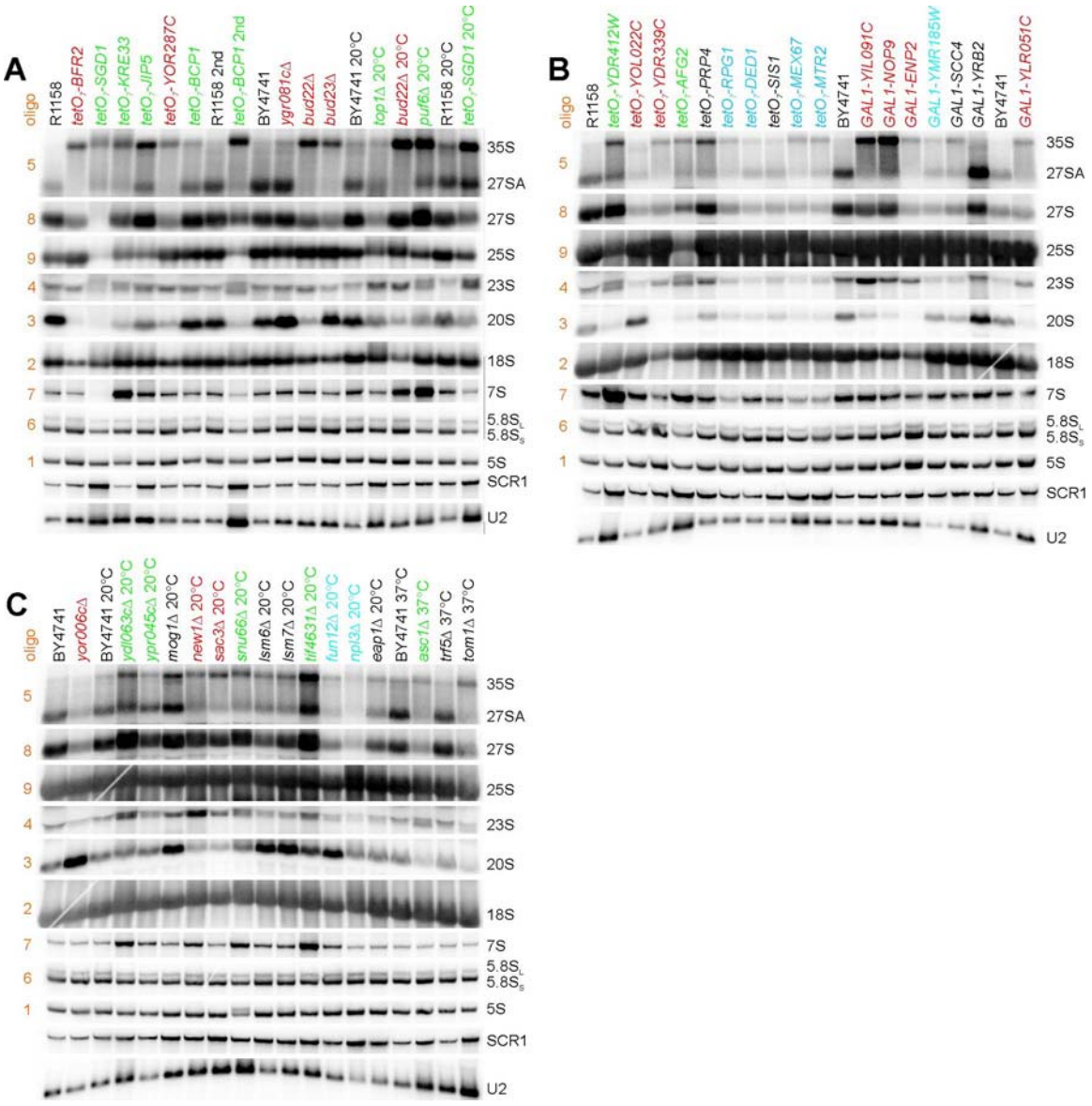


Figure 5.9 Northern blots of pre-rRNA and rRNA species in each mutant.
Each strain is color-labeled as in Figure 5.4. The oligo numbers on the left side of each blot are indicated on the rDNA repeat in Figure 5.7A. The detected RNA species are on the right side of each blot. SCR1 and U2 are internal controls.

BUD23 was also implicated in 35S processing (Figure 5.9A). Additionally, we observed a slight accumulation of the 20S pre-rRNA in *bud23Δ*, suggesting either a processing defect for 20S pre-rRNA or an export defect for 43S pre-ribosome caused accumulation of 20S in the nucleus.

Either depletion of Prp4p or deletion of *SAC3* or *NEW1* resulted in accumulation of 35S and reduction of 27SA and 20S, suggesting the processing defects at sites A_0-A_2, but no dramatic reduction of 18S (Figure 5.9B-C). Sac3p forms a stable complex with Thp1p *in vivo* and mutation of either of these genes leads to mRNA export defects (Fischer et al., 2002). Deletion of *SAC3* might indirectly affect pre-rRNA processing by inhibiting exports of mRNAs for r-proteins, causing the lack of r-proteins required for ribosome assembly. However, the ribosome biogenesis defect for *sac3Δ* was only observed at 20°C, not at 30°C or 37°C (Figure 5.10), whereas the mRNA export in *sac3Δ* was prevented at both 23°C and 37°C (Fischer et al., 2002). In addition, deletion of *THP1* did not show obvious ribosome biogenesis defects in the polysome profile analysis (data not shown), further suggesting that mRNA export defects might not explain the ribosome biogenesis defects observed for *sac3Δ*. *NEW1* encodes an ATP binding protein and is involved in prion formation (Osherovich and Weissman, 2001). Deletion of *NEW1* showed an increase of 23S and a minor increase of 7S besides accumulation of 35S and reduction of 27S and 20S (Figure 5.9C). Most of the New1 proteins distributed in the polysome fractions in the sucrose density gradient (Figure 5.6F), consistent with its cytoplasmic localization and the observation of its interaction with translation factors in a large-scale affinity purification assay (Gavin et al., 2006). *NEW1* might thus be indirectly linked to ribosome biogenesis by affecting translation.

PRP4 is a component of the U4/U6.U5 snRNP complex involved in splicing (Stevens and Abelson, 1999). We observed a 35S processing defect after depletion of

PRP4 (Figure 5.9B). A large scale affinity purification study has shown that Prp4p interacts with many SSU processome components such as Mpp10p, Nop58p, Rrp9p, et al (Krogan et al., 2006). These results therefore suggested that *PRP4* might have dual functions in both splicing and ribosome biogenesis.

5.3.5.2 Genes involved in 20S pre-rRNA processing

The 20S pre-rRNA is exported from the nucleolus to the cytoplasm as part of the 43S pre-ribosome where it is cleaved at site D to form the mature 18S rRNA (Figure 5.7B). We observed accumulations of 20S upon deletion of the genes *YOR006C*, *YGR081C*, *MOG1*, *FUN12*, *LSM6*, or *LSM7*, or depletion of Yol022cp or Yrb2p (Figure 5.8), indicating either defective cleavage at site D or insufficient 43S particle export from the nucleus to the cytoplasm. Depletion of Yol022cp or deletion of *YOR006C* also led to reduction of 27S pre-rRNA without accumulation of 35S pre-rRNA (Figure 5.9B-C). Deletion of *YGR081C* slightly affected the 35S processing, whereas 27SA and 27S species were largely unaffected (Figure 5.9A). In the co-sedimentation assay, Yol022cp and Yor006cp did not show obvious association with the 40S subunits; however, Ygr081cp co-sedimented with the 40S fractions suggesting that *YGR081C* might play a direct role in either 20S processing or small subunit export (Figure 5.6A). Yol022cp and Yor006cp might therefore interact with the small subunit transiently or affect 20S processing indirectly. Recently, *YGR081C* was confirmed to participate in ITS1 processing (Bax et al., 2006).

All the r-proteins and most of the trans-acting proteins in ribosome biogenesis must be imported into the nucleus and/or nucleolus. Mog1p is a conserved protein involved in nuclear protein import (Oki and Nishimoto, 1998). The pre-rRNA processing defects observed for *mog1Δ* (Figure 5.9C) might therefore stem from insufficient nuclear import of the necessary proteins required for ribosome biogenesis. Fun12p is a conserved

translation initiation factor that promotes Met-tRNAiMet binding to the ribosomes (Choi et al., 1998). Deletion of *FUN12* reduced the levels of 27S and accumulated 20S (Figure 5.9C). Fun12p not only interacted with many r-proteins, but also interacted with many ribosome biogenesis factors in large-scale affinity purification studies (Gavin et al., 2006; Krogan et al., 2004). In addition, we observed the polysome profile of *fun12Δ* distinct from the deletion of other translation initiation factor genes such as *TIF34* and *TIF35* (Figure 5.4C). The evidence therefore suggests that Fun12p might be a novel factor involved in both ribosome biogenesis and translation initiation. Lsm6p and Lsm7p are components of Lsm1p-7p and Lsm2p-8p complexes involved in the mRNA decay and nuclear RNA processing, respectively (Beggs, 2005; Tharun et al., 2000). Depletion of the essential Lsm2-5 or Lsm8 proteins led to the delay of pre-rRNA processing and the accumulation of aberrant processing intermediates (Kufel et al., 2003). We observed that deletion of the non-essential Lsm6p or Lsm7p led to the accumulation of 35S and 20S (Figure 5.9C), supporting the notion that the Lsm complex affects pre-rRNA processing. Yrb2p is involved in the nuclear protein export (Taura et al., 1998) and particularly export of the ribosome small subunit (Moy and Silver, 2002). The accumulation of 20S pre-rRNA might reflect the accumulation of pre-40S in the nucleus (Figure 5.9B). Additionally, the accumulation of 27SA but not 27S implicates a processing defect from 27SA to 27SB (Figure 5.9B).

5.3.5.3 Genes required for 27S processing

Defective processing of 27S and/or 7S often leads to the insufficient synthesis of 60S large subunit ribosomes, frequently accompanied with delayed 35S processing and reduced 20S levels (Venema and Tollervey, 1999). We observed one set of mutants with defects in the 27S and/or 7S processing, including *tetO7-SGD1*, *tetO7-JIP5*, *tetO7-BCP1*,

top1Δ, *ydl063cΔ*, *ypr045cΔ*, *asc1Δ*, *tetO7-YDR412W*, *tetO7-AFG2*, *tetO7-KRE33*, *puf6Δ* and *tif4631Δ*, most of which also accumulated 35S (Figure 5.8).

Sgd1p was previously shown to interact with Plc1p and was involved in osmoregulation (Lin et al., 2002). Depletion of *SGD1* led to the reduction of 27S, 7S, 25S and 5.8S (Figure 5.9A). We also observed Sgd1p to co-sediment with 60S fractions in the sucrose density gradient (Figure 5.6A), suggesting that *SGD1* is most likely directly involved in ribosome biogenesis.

We observed *JIP5*, *BCP1*, *TOP1*, *YDL063C*, and *YPR045C* specifically affected the upstream processing of 35S and/or 27S, whereas *YDR412W*, *PUF6*, and *TIF4631* strongly affected 7S processing, as well as 27S processing (Figure 5.9A-C). Depletion of *KRE33* led to the accumulation of 35S and 7S without the accumulation of 27S (Figure 5.9A). Interestingly, depletion of *KRE33* also caused underaccumulation of the 60S subunits relative to the 40S subunits without inducing halfmers in a polysome profile analysis (Figure 5.4B). Depletion of *AFG2* showed a large reduction of both 27S and 25S and a slight increase of 7S (Figure 5.9B); its role in ribosome biogenesis was recently independently confirmed during the course of the work (Pertschy et al., 2007).

ASC1 contains an intron that encodes a C/D box small nucleolar RNA U24 required for 2'-O-methylation of 25S at C1437, C1449, and C1450 (Kiss-Laszlo et al., 1996), whereas Asc1 protein has been shown to be a component of the 40S subunit (Gerbasi et al., 2004). Previous studies have shown that deletion of the intron and the first exon of *ASC1* did not cause significant pre-rRNA processing defects (Kiss-Laszlo et al., 1996). However, we observed temperature-dependent reductions of 27S, 20S, and 25S upon deletion of both intron and exons of *ASC1* when cultured at 37°C (Figure 5.9C). Additionally, the *asc1Δ* mutant showed a typical 60S subunit biogenesis defect in the polysome profile analysis when cultured at 37°C (Figure 5.4B). Whether the observed

ribosome biogenesis defects are due to the deletion of intron or the gene for Asc1 protein remains to be determined.

5.3.5.4 SNU66 is involved in processing the 5S rRNA precursor

The 5S rRNA precursor is transcribed by RNA polymerase III and subsequently processed by 3' exonuclease Rna82p/Rex1p/Rnh70p (Figure 5.7B) (Piper et al., 1983; van Hoof et al., 2000). In addition to the processing defects for 35S, 27S, and 7S upon deletion of *SNU66*, we observed an inefficient processing of 5S rRNA precursor (Figure 5.9C), the only such defect observed among the 44 mutants tested. Snu66p is a known component of the U4/U6.U5 snRNP complex involved in pre-mRNA splicing (Stevens and Abelson, 1999). Splicing defects might indirectly affect ribosome biogenesis since 99 out of 137 genes for r-proteins contain introns (Planta and Mager, 1998). However, the unique processing defect for 5S rRNA precursor was not observed upon depletion of Prp4p, another component of U4/U6.U5 snRNP (Figure 5.9B). These results thus suggest that *SNU66* might be involved in both splicing and 5S rRNA biogenesis.

5.3.5.5 Other genes affecting pre-rRNA processing

Finally, we identified a class of mutants exhibiting a slight accumulation of 35S accompanied by dramatic reduction of 27S and 20S, including *tetO7-RPG1*, *tetO7-DED1*, *tetO7-SIS1*, *tetO7-MEX67*, *tetO7-MTR2*, *tom1Δ*, *npl3Δ*, *GAL1-YMR185W*, and *GAL1-SCC4* (Figure 5.9A-C). These mutants also showed imbalanced 40S to 60S subunits ratios in polysome profile analyses when compared to wild type strains (Figure 5.4C and Figure 5.10). Rpg1p and Ded1p are translation initiation factors (Iost et al., 1999; Valasek et al., 2003), and Sis1p is a chaperone belonging to the Hsp40p family (Fan et al., 2004). Mex67p, Mtr2p, Tom1p, and Npl3p are known factors involved in mRNA export (Duncan et al., 2000; Lee et al., 1996; Santos-Rosa et al., 1998). The pre-

rRNA processing defects observed for the mutants are possible secondary consequences of the mRNA export and translation defects. Scc4p forms a complex with Scc2p, facilitating the loading of cohesin complexes onto the chromosomes (Ciosk et al., 2000) and Ymr185wp is an essential protein with unknown function. Interestingly, Scc4p was previously shown to interact with several ribosome biogenesis factors, including Nop1p, Rrp1p, and Utp22p (Krogan et al., 2006).

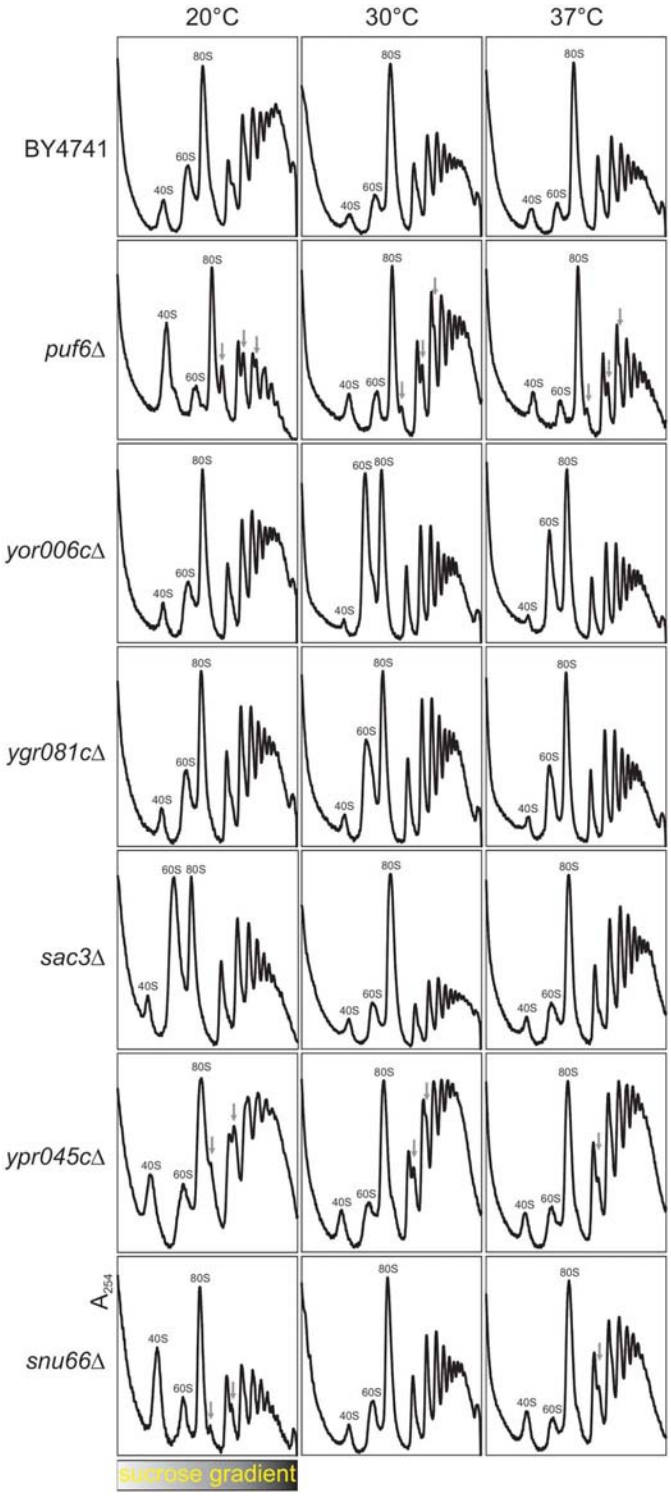


Figure 5.10 Polysome profiles of several mutants cultured at different temperature conditions.

5.3.6 New genes involved in ribosomal subunit export

Ribosome export to the cytoplasm is dependent on the nuclear pore complex (NPC), the RanGTPase system, an export receptor such as Crm1p, and an export adaptor such as Nmd3p, a trans-acting factor associated with late pre-60S particles (Johnson et al., 2002; Zemp and Kutay, 2007). However, defective pre-rRNA processing and/or ribosome assembly may also lead to inefficient transport of ribosomes to the cytoplasm (Milkereit et al., 2001). To test whether the ribosome biogenesis candidates affect ribosome transport, we assayed ribosome export in the mutants by using Rps2-GFP and Rpl25-GFP as reporters for the small and large ribosomal subunits, respectively (Grandi et al., 2002; Hurt et al., 1999), while monitoring the nucleolus with Sik1-mRFP (Gautier et al., 1997; Huh et al., 2003). In wild type control strains cultured at various conditions, both small and large ribosomal subunits localized primarily in the cytoplasm (Figure 5.11). Upon depletion of Yrb2p, a known factor involved in small subunit export (Moy and Silver, 2002), ribosomal small subunits accumulated in the nucleus while the large subunits were unaffected (Figure 5.12).

In mutants defective in the synthesis of small subunits, such as *tetO7-BFR2*, *bud22Δ*, *bud23Δ*, *tetO7-YDR339C*, *ygr081cΔ*, *GAL1-ENP2*, and *GAL1-NOP9*, we observed significant accumulation of small subunits in the nucleus, whereas the large subunits localized in the cytoplasm similarly to control strains (Figure 5.12). Quantitative image analyses indicated that in *tetO7-BFR2* and *tetO7-YDR339C*, the small subunits were enriched in the nucleolus as well, implicating their involvement in early stages of ribosome biogenesis (Figure 5.12).

In mutants with defective synthesis of large ribosomal subunits, such as *tetO7-AFG2*, *tetO7-BCP1*, *tetO7-KRE33*, *puf6Δ*, *tetO7-SGD1*, *tetO7-YDR412W*, and *tif4631Δ*, strong accumulation of the large ribosomal subunits in the nucleolus and nucleus was

observed, but not of the small subunits (Figure 5.13). In *tetO7-KRE33* and *puf6Δ*, the large ribosomal subunits strongly accumulated in the nucleolus (Figure 5.13). Interestingly, deletion of *LSM6* or *LSM7* inhibited the transport of ribosomal large subunits to the cytoplasm, but not the small subunits (Figure 5.13). Tom1p is a HECT-domain ubiquitin ligase involved in many processes including transcriptional regulation (Saleh et al., 1998), mRNA export (Duncan et al., 2000), and maintenance of nuclear structure (Utsugi et al., 1999). The *tom1Δ* mutant showed strong accumulation of the ribosomal small subunits in the nucleus at 37°C (Figure 5.12), whereas about 50% of cells showed nuclear accumulation of the large ribosomal subunits (Figure 5.13). In total, we identified 17 new genes that affected export of either the ribosomal small or large subunits.

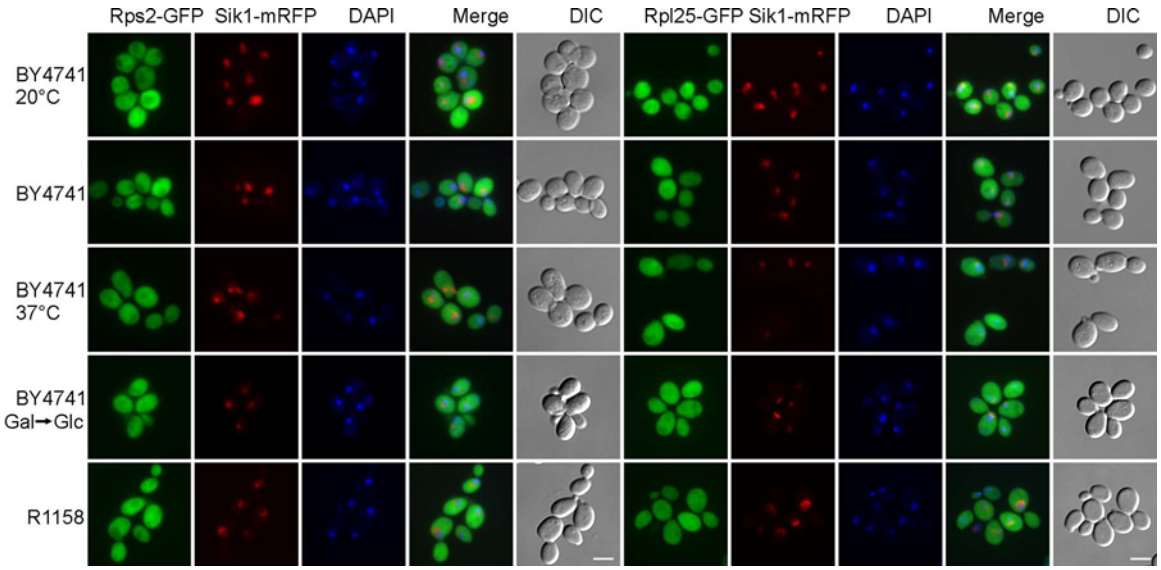


Figure 5.11 Ribosomal subunits export assays for control strains.
Strains were cultured at 30°C unless otherwise indicated. Rps2-GFP and Rpl25-GFP were used as reporters for ribosomal small and large subunits, respectively. Sik1-RFP was used as the nucleolus marker. DAPI was used to stain DNA for visualizing the nucleus. The white scale bar represents 5um.

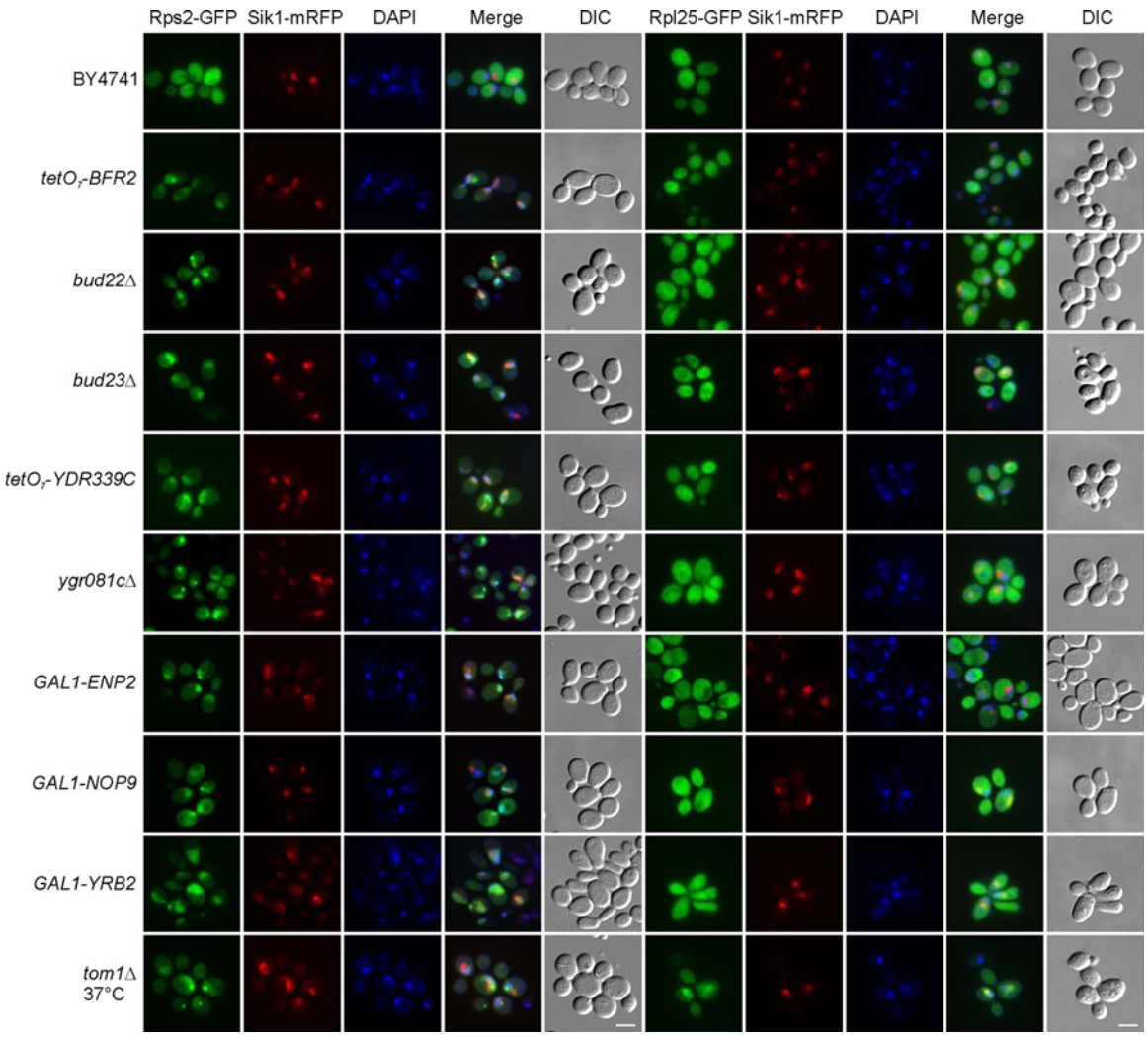


Figure 5.12 Mutants with ribosomal small subunit export defects.
One representative control is shown at the first row. Strains were cultured at 30°C unless otherwise indicated. Rps2-GFP and Rpl25-GFP were used as reporters for ribosomal small and large subunits, respectively. Sik1-RFP was used as the nucleolus marker. DAPI was used to stain DNA for visualizing the nucleus. The white scale bar represents 5um.

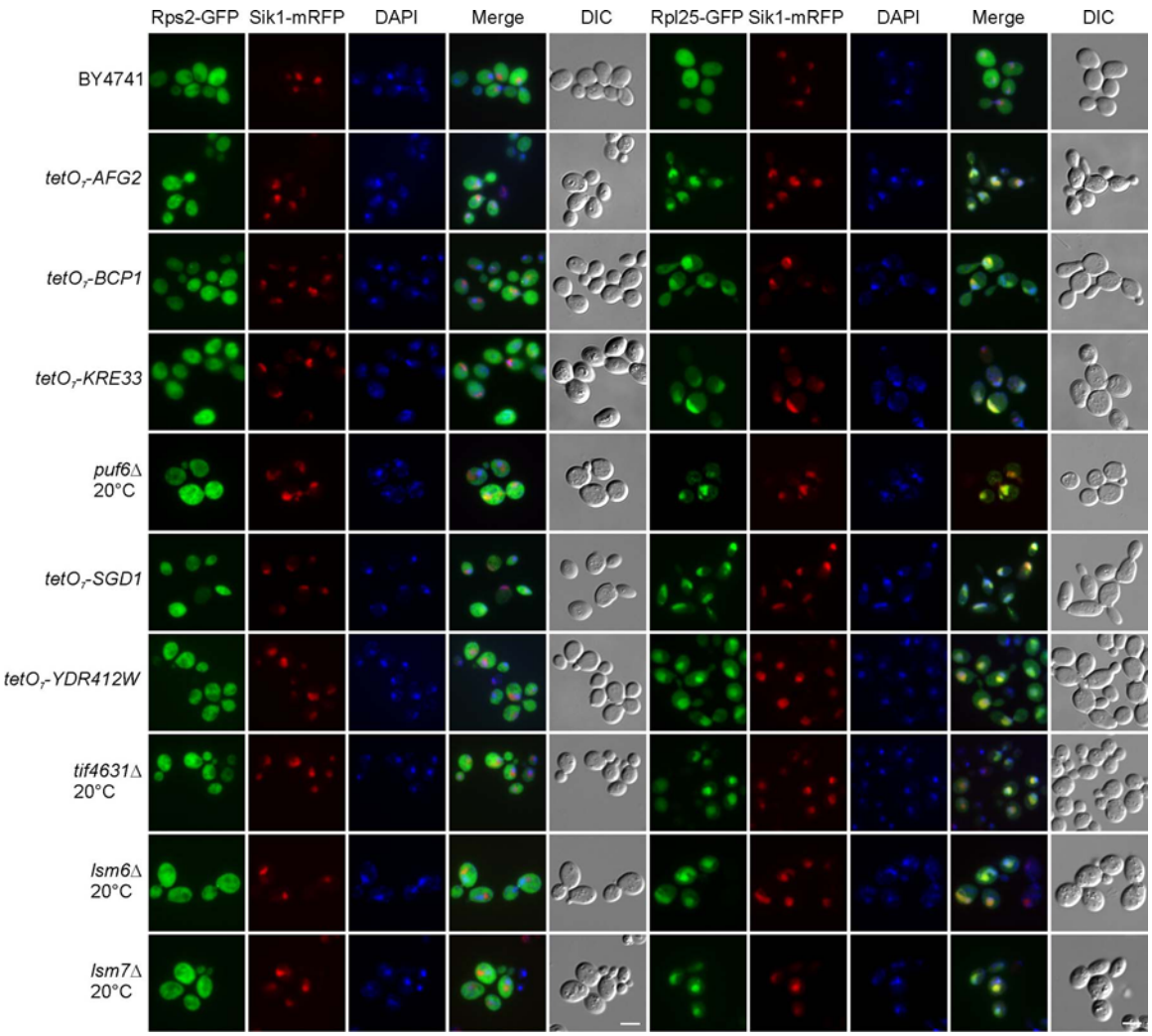


Figure 5.13 Mutants with ribosomal large subunit export defects.
One representative control is shown at the first row. Strains were cultured at 30°C unless otherwise indicated. Rps2-GFP and Rpl25-GFP were used as reporters for ribosomal small and large subunits, respectively. Sik1-RFP was used as the nucleolus marker. DAPI was used to stain DNA for visualizing the nucleus. The white scale bar represents 5um.

5.4 DISCUSSION

We applied computational predictions using a functional gene network, followed by experimental validations to discover at least 25 previously unreported ribosome biogenesis genes (Table 5.3). Network-based prediction therefore proved to be a powerful approach for identifying new genes in a pathway, even in such a well-studied cellular process as ribosome biogenesis, with ~30% of the tested genes in polysome profile analyses shown to participate in this pathway. Although considerable effort has been spent predicting and validating gene functions from diverse functional genomics and proteomics data (Peng et al., 2003; Sharan et al., 2007), to our knowledge, this is the first systematic large-scale experimental test of gene network-based predictions for a conserved fundamental cellular process.

Table 5.3 Summary of the evidence for involvement in ribosome biogenesis

ORF	Gene	# links to seed genes	Network evidence	Mutant growth	Polysome profile defect	Co-sedimentation	Pre-RNA processing defect	Ribosome export defect
YPR045C	MNI2	5	MS	Slow	60S	60S	35S	No
YGR162W	TIF4631	22	MS, CX, LC	Slow	60S	Across gradient	35S, 27S, 7S, 20S	60S
YOR308C	SNU66	8	MS, CC, LC	Slow at 20C	60S	40S	35S, 27S, 5S	50% cells 60S
YDL063C	-	5	MS, CC, YH, CX	Slow	60S	Free	35S, 27S	No
YDR412W	RRP17	14	CX, MS, YH	Essential	60S	Free	35S, 7S	60S
YPR169W	JIP5	19	CX, MS	Essential	60S	Free, 60S	35S, 27S	50% cells 60S
YOL006C	TOP1	7	CC, MS, LC, CX	Slow	60S	Across fraction	35S, 27S	No
YNL132W	KRE33	77	MS, CX, LC	Essential	60S	Free, 60S	35S, 7S	60S
YDR496C	PUF6	94	CX, MS, LC	Slow at 20C	60S	60S	35S, 27S, 7S	60S
YLR336C	SGD1	31	CX, MS	Essential	60S	60S	35S	60S
YLR397C	AFG2	7	CX, MS, CC	Essential	60S	-	35S, 7S	60S
YDR361C	BCP1	19	CX	Essential	60S	Free, 60S	35S	60S
YJL010C	NOP9	56	CX, LC	Essential	40S	40S, Polysome	35S	40S
YOR287C	-	40	CX, MS	Essential	40S	-	35S	No
YDR339C	FCF1	13	CX	Essential	40S	-	35S	40S
YMR014W	BUD22	37	CX, MS	Slow	40S	80/90S, Polysome	35S	40S
YCR047C	BUD23	7	MS, CX	Slow	40S	40S	35S, 20S	40S
YLR051C	FCF2	13	CX	Essential	40S	-	35S	-
YGR145W	ENP2	91	CX, MS, LC, RS	Essential	40S	-	35S	40S
YDR299W	BFR2	71	CX, MS, LC	Essential	40S	40S, 80/90S	35S	40S
YIL091C	-	12	CX, MS	Essential	40S	40S	35S	No
YOL022C	-	30	CX	Essential	40S	Free	20S	No
YOR006C	-	2	CX	Slow at 30C	40S	Free	20S	No
YGR081C	SLX9	14	MS, CX, GT	Slow at 30C	40S	40S	20S	40S
YDR159W	SAC3	1	LC	Slow	40S	40S, 80/90S	35S	No
YPL226W	NEW1	8	CX, MS	Slow at 20C	40S	Across gradient	35S	No
YJR074W	MOG1	3	CC, GT, MS, LC, YH	Slow	Minor	Free	35S, 27S, 20S	No
YAL035W	FUN12	40	MS, GN, CX	Slow	40S	Polysome	20S	No
YPR178W	PRP4	11	MS, LC, CC, YH	Essential	Minor	Free, 40S	35S	No
YDR378C	LSM6	7	MS, LC, CC, YH, TS	Slow at 20C	Minor	-	35S, 20S	50% cells 60S
YNL147W	LSM7	7	MS, LC, CC, YH, TS,	Slow at 20C	Minor	Polysome	35S, 20S	50% cells 60S

CC: co-citation, CX: co-expression, GN: gene neighbor, GT: genetic interaction, LC: literature curate protein-protein interaction, MS: mass spectrometry analysis of affinity complex, PG: phylogenetic profile, RS: Rosetta Stone protein (gene fusion), TS: protein tertiary structure based inferred protein-protein interaction, YH: high-throughput yeast two hybrid.

As expected, many genes for ribosome biogenesis are essential. However, a large number of non-essential genes are clearly involved in ribosome biogenesis, some of which showed extremely strong conditional phenotypes (Figure 5.10). The polysome profile of *puf6Δ* showed strong 60S biogenesis defects at 20°C, while only minor defects were observed at the optimal temperature 30°C. Deletion of *SAC3* or *SNU66* exhibited no obvious defects for both ribosomal subunits at 30°C, but defects were observed for small subunits or large subunits respectively at 20°C. Interestingly, the polysome profile of *yor006cΔ* showed 40S biogenesis defects at 30°C, but no defects at 20°C. Several non-essential genes, including *YIL096C*, *YCR016W*, *YJL122W*, *YNL022C*, *BUD20*, and *NOP13*, form a tight cluster with known ribosome biogenesis genes in the gene network and their encoded proteins co-sedimented with either 40S or 60S fractions, supporting them as components of pre-ribosomes (data not shown). However, deletion mutants for those genes did not show growth defects at 20°C, 30°C, or 37°C (Figure 5.2), nor were polysome profiles of the deletion mutants different from wild type cells (data not shown). Yjl122p (Alb1p) was recently confirmed to interact with the known ribosome biogenesis factor Arx1p directly, although the deletion mutant had no observable phenotype (Lebreton et al., 2006). It is therefore still likely that these genes participate in ribosome biogenesis, but that we failed to identify a conditional phenotype or that these genes are functionally redundant with other genes. In the latter case, synthetic lethal assays might prove a useful strategy for deciphering the genes' functions. Indeed, we observed one such example: mutants with either deletion of *TRF5* or depletion of *PAP2* did not exhibit defects in polysome profile analyses at 30°C; however, mutant with both deletion of *TRF5* and depletion of *PAP2* exhibited strong 60S biogenesis defects in polysome profile analysis (Figure 5.14), suggesting that *TRF5* and its paralog *PAP2* were involved in

ribosome biogenesis. Thus, many of the remaining non-essential mutants without conditional phenotypes may still be involved in ribosome biogenesis.

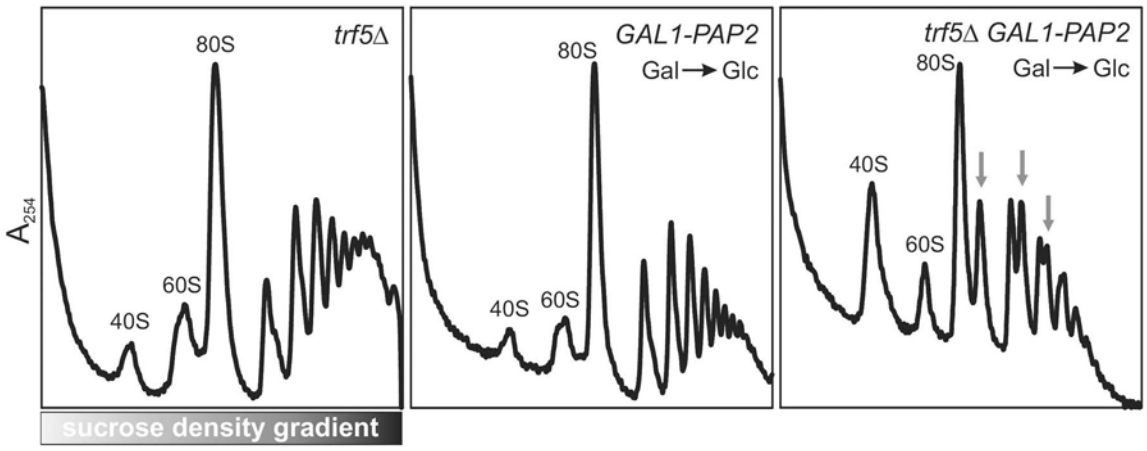


Figure 5.14 Synthetic ribosome biogenesis defects observed in a double mutant *trf5Δ GAL1-PAP2*.
trf5Δ mutant was cultured in YPD. *GAL1-PAP2* and *trf5Δ GAL1-PAP2* were first cultured in YPGal, then diluted into YPD and cultured to early logarithm phase. Gray arrows indicate halfmer polysomes.

Gene network-based predictions based upon binary associations between genes intrinsically help to identify genes that participate in multiple cellular processes. Correspondingly, several genes we identified have been reported to have other functions: for example, *PUF6* is required for *ASH1* mRNA asymmetric localization (Gu et al., 2004). Sgd1p interacts with Plc1p and is involved in osmoregulation (Lin et al., 2002). Sac3p localizes at the nuclear pore and is involved in mRNA export (Fischer et al., 2002); Snu66p is a component of the tri-snRNP involved in mRNA splicing (Stevens and Abelson, 1999). Recent study showed that Mtr2p, known as a mRNA export receptor (Santos-Rosa et al., 1998) is directly involved in ribosomal large subunit export (Yao et al., 2007). Similarly, we identified *SAC3* as a ribosome biogenesis gene based on polysome profile and Northern blot analyses of the deletion mutant (Figure 5.4D, 5.9C). Additionally, Sac3p co-sedimented with 40S fractions, suggesting its possible association with ribosomes (Figure 5.6A). It is also known that Sac3p can mediate protein export (Jones et al., 2000); However, we did not observe export defects for either ribosomal subunit in the *sac3Δ* mutant (data not shown). Recently, the spicing factor Prp43p was confirmed to be a ribosome biogenesis factor by several groups, suggesting coordination of ribosome biogenesis and mRNA splicing (Combs et al., 2006; Lebaron et al., 2005; Leeds et al., 2006). Snu13p, a component of the tri-snRNP involved in mRNA splicing, is also part of the U3 snoRNP and involved in pre-rRNA processing (Dobbyn and O'Keefe, 2004). Finally, we observed that another component of the tri-snRNP, Snu66p, not only delayed 35S processing, but also affected processing of the 5S rRNA precursor (Figure 5.9C), further supporting a connection between ribosome biogenesis and mRNA splicing.

The success of our network prediction approach relies not only on the improved computational algorithm, but also on the quality and coverage of the genomic scale data sets. However, current large-scale data are still far from complete (Reguly et al., 2006).

Hence, new ribosome biogenesis genes will certainly continue to emerge in the future. Additionally, the network-guided prediction approach might be also useful to identify gene functions for many other cellular processes.

Chapter 6: Summary

Protein is one of the major biological macromolecules in the cell. To understand cells at a molecular level as well as a system level, we need to identify all the proteins in the cell (proteome), elucidate their interactions with each other and interplay with other molecules (interactome), and determine their functions in the cell. In particular, interactions among proteins are central to their biological functions. During the past decade, protein-protein interactions are extensively studied in a large-scale and high-throughput fashion by using a variety of approaches, especially in the model organism *Saccharomyces cerevisiae* (Shoemaker and Panchenko, 2007a; Shoemaker and Panchenko, 2007b). Many of the approaches have their own merits as well as limitations, thereby complementing to each other in protein-protein interaction studies. The goal to study protein-protein interactions is to determine the functions of the proteins and understand how they work at a system level.

In this dissertation, I first explored the possibility of applying FRET to study protein-protein interactions in a large-scale. By using fluorescence proteins ECFP as the donor molecule and EYFP as the acceptor molecule, FRET signal was observed for EYFP-ECFP fusion protein expressed in E. coli by using a fluorometer. However, emission fluorescent signal of EYFP was also observed by using the excitation wavelength for FRET due to the overlap of the excitation spectrum between ECFP and EYFP, thereby confounding the FRET analysis. As an example, the known interacting partners WW domain and a peptide with PPXY motif were fused with ECFP and EYFP, respectively, and interactions between WW domain and this peptide were monitored by exciting ECFP. The quenching of fluorescence donor ECFP was observed but rather weak. In this case, the efficiency of expressing different fusion proteins differs

significantly, with large quantities of fusion proteins of the peptide and EYFP and little expression of fusion proteins of WW domain and ECFP. The difference of expression levels of the donor and acceptor will prevent detecting FRET signals effectively. In conclusion, it is not practical to monitor protein-protein interactions in a large-scale by using ECFP and EYFP as a FRET pair. Development of new fluorescent proteins for FRET assay would be necessary.

Tandem affinity purification by using TAP tag was a new approach to identifying interaction partners by then. This method was employed to purify several transcription factor complexes in yeast: histone deacetylase complex B (HDB), Hap2/3/5 complex, and Swi6-containing complexes. Affinity purified samples were analyzed by using a shotgun mass spectrometry approach, and each identified protein was ranked based on the protein score from data analysis by SEQUEST. By comparing the rankings of proteins identified from purified samples to the rankings of proteins identified from the whole cell lysate control sample, specific proteins in the purified protein complex were identified. Several novel components in the Rpd3-containing complex (Dep1p, Ybr095cp, Ymr263cp, Ume1p, Ymr075wp, and Fun19p) were identified, and most of which were later confirmed and characterized (Carrozza et al., 2005) except Fun19p. Whi5p was identified as a component in SBF complex (Swi6-Swi4), and later characterized as a negative regulator for G1 to S transition in cell cycle (Costanzo et al., 2004; de Bruin et al., 2004). Interestingly, Gln3p was identified as a component in Hap2/3/5 complex, implicating a crosstalk between nitrogen metabolism and carbon metabolism. However, no obvious synthetic genetic interaction was observed between *GLN3* and *HAP3*. Gene expression profiling by DNA microarray for single and double deletion mutants showed that a fraction of genes were co-regulated by Gln3p and Hap complex. Interestingly, from the gene expression data, it seems that both Gln3p and Hap complex function not only under

nutrient starvation condition but also under rich medium condition. These affinity purification examples demonstrate the power of this approach. To further extend and improve this approach, a simple method by combining one-step affinity purification with shotgun mass spectrometry was developed and proved useful for identification of Whi5p in SBF complex.

In order to analyze protein complexes in their native states, a new method used to map protein complexes was proposed, which measures protein correlation profiles across the sucrose density gradient by state-of-the-art shotgun mass spectrometry. In two examples for yeast and Hela cells, several known complexes were identified, demonstrating the utility of this method in analyzing protein associations in the cell. This method can be further improved in terms of applying other protein complex separation techniques (such as isoelectric focusing and native gel electrophoresis) and developing statistic models to analyze the data.

Finally, to demonstrate the utility of protein-protein interaction information, a yeast gene interaction network reconstructed from diverse large-scale interaction data sets was used to predict ribosome biogenesis genes based on gene connections in the network. The predicted genes were further experimentally analyzed by several methods: growth assays, polysome profile analyses, protein co-sedimentation analyses, Northern blot analyses, and ribosomal subunit export assays. About 30 genes were confirmed to be involved in ribosome biogenesis, of which 25 genes were previously unreported, underscoring the value of protein-protein interaction data and gene interaction network.

Bibliography

Adams, C. C., Jakovljevic, J., Roman, J., Harnpicharnchai, P., and Woolford, J. L., Jr. (2002). Saccharomyces cerevisiae nucleolar protein Nop7p is necessary for biogenesis of 60S ribosomal subunits. Rna *8*, 150-165.

Alland, L., David, G., Shen-Li, H., Potes, J., Muhle, R., Lee, H. C., Hou, H., Jr., Chen, K., and DePinho, R. A. (2002). Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. Mol Cell Biol *22*, 2743-2750.

Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. Nature *426*, 570-574.

Arrigo, A. P., Tanaka, K., Goldberg, A. L., and Welch, W. J. (1988). Identity of the 19S 'prosome' particle with the large multifunctional protease complex of mammalian cells (the proteasome). Nature *331*, 192-194.

Bassler, J., Grandi, P., Gadal, O., Lessmann, T., Petfalski, E., Tollervey, D., Lechner, J., and Hurt, E. (2001). Identification of a 60S preribosomal particle that is closely linked to nuclear export. Mol Cell *8*, 517-529.

Bastiaens, P. I., and Jovin, T. M. (1996). Microspectroscopic imaging tracks the intracellular processing of a signal transduction protein: fluorescent-labeled protein kinase C beta I. Proc Natl Acad Sci U S A *93*, 8407-8412.

Bax, R., Raue, H. A., and Vos, J. C. (2006). Slx9p facilitates efficient ITS1 processing of pre-rRNA in Saccharomyces cerevisiae. Rna *12*, 2005-2013.

Beck, T., and Hall, M. N. (1999). The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. Nature *402*, 689-692.

Beggs, J. D. (2005). Lsm proteins and RNA processing. Biochem Soc Trans *33*, 433-438.

Bernstein, K. A., Gallagher, J. E., Mitchell, B. M., Granneman, S., and Baserga, S. J. (2004). The small-subunit processome is a ribosome assembly intermediate. Eukaryot Cell *3*, 1619-1626.

Bleichert, F., Granneman, S., Osheim, Y. N., Beyer, A. L., and Baserga, S. J. (2006). The PINc domain protein Utp24, a putative nuclease, is required for the early cleavage steps in 18S rRNA maturation. Proc Natl Acad Sci U S A *103*, 9464-9469.

Boozer, C., Kim, G., Cong, S., Guan, H., and Londergan, T. (2006). Looking towards label-free biomolecular interaction analysis in a high-throughput format: a review of new surface plasmon resonance technologies. Curr Opin Biotechnol *17*, 400-405.

Borggrefe, T., Davis, R., Bareket-Samish, A., and Kornberg, R. D. (2001). Quantitation of the RNA polymerase II transcription machinery in yeast. J Biol Chem *276*, 47150-47153.

Brajenovic, M., Joberty, G., Kuster, B., Bouwmeester, T., and Drewes, G. (2004). Comprehensive proteomic analysis of human Par protein complexes reveals an interconnected protein network. J Biol Chem *279*, 12804-12811.

Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N.*, et al.* (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature *433*, 531-537.

Camacho-Carvajal, M. M., Wollscheid, B., Aebersold, R., Steimle, V., and Schamel, W. W. (2004). Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. Mol Cell Proteomics *3*, 176-182.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Carvalho, J., and Zheng, X. F. (2003). Domains of Gln3p interacting with karyopherins, Ure2p, and the target of rapamycin protein. J Biol Chem *278*, 16878-16886.

Chen, Y. I., Maika, S. D., and Stevens, S. W. (2006). Epitope tagging of proteins at the native chromosomal loci of genes in mice and in cultured vertebrate cells. J Mol Biol *361*, 412-419.

Choi, S. K., Lee, J. H., Zoll, W. L., Merrick, W. C., and Dever, T. E. (1998). Promotion of met-tRNAiMet binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. Science *280*, 1757-1760.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. Molecular Cell *5*, 243-254.

Combs, D. J., Nagel, R. J., Ares, M., Jr., and Stevens, S. W. (2006). Prp43p is a DEAH-box spliceosome disassembly factor essential for ribosome biogenesis. Mol Cell Biol *26*, 523-534.

Costanzo, M., Nishikawa, J. L., Tang, X., Millman, J. S., Schub, O., Breitkreuz, K., Dewar, D., Rupes, I., Andrews, B., and Tyers, M. (2004). CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell *117*, 899-913.

Courchesne, W. E., and Magasanik, B. (1988). Regulation of nitrogen assimilation in Saccharomyces cerevisiae: roles of the URE2 and GLN3 genes. J Bacteriol *170*, 708-713.

Day, R. N., Periasamy, A., and Schaufele, F. (2001). Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus. Methods *25*, 4-18.

de Bruin, R. A., McDonald, W. H., Kalashnikova, T. I., Yates, J., 3rd, and Wittenberg, C. (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell *117*, 887-898.

Deeds, E. J., Ashenberg, O., and Shakhnovich, E. I. (2006). A simple physical model for scaling in protein-protein interaction networks. Proc Natl Acad Sci U S A *103*, 311-316.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680-686.

Devin, A. B., and Koltovaya, N. A. (1987). Genetic modification of the spontaneous rho− mutability in Saccharomyces cerevisiaee. Current Genetics *11*, 411-413.

Dobbyn, H. C., and O'Keefe, R. T. (2004). Analysis of Snu13p mutations reveals differential interactions with the U4 snRNA and U3 snoRNA. Rna *10*, 308-320.

Dragon, F., Gallagher, J. E., Compagnone-Post, P. A., Mitchell, B. M., Porwancher, K. A., Wehner, K. A., Wormsley, S., Settlage, R. E., Shabanowitz, J., Osheim, Y.*, et al.* (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature *417*, 967-970.

Dudkina, N. V., Eubel, H., Keegstra, W., Boekema, E. J., and Braun, H. P. (2005). Structure of a mitochondrial supercomplex formed by respiratory-chain complexes I and III. Proc Natl Acad Sci U S A *102*, 3225-3229.

Duncan, K., Umen, J. G., and Guthrie, C. (2000). A putative ubiquitin ligase required for efficient mRNA export differentially affects hnRNP transport. Current Biology *10*, 687-696.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.

Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. Nature *402*, 86-90.

Ewing, R. M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M. D., O'Connor, L., Li, M.*, et al.* (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol *3*, 89.

Fan, C.-Y., Lee, S., Ren, H.-Y., and Cyr, D. M. (2004). Exchangeable Chaperone Modules Contribute to Specification of Type I and Type II Hsp40 Cellular Function. Mol Biol Cell *15*, 761-773.

Fatica, A., and Tollervey, D. (2002). Making ribosomes. Curr Opin Cell Biol *14*, 313-318.

Fearnley, I. M., Carroll, J., Shannon, R. J., Runswick, M. J., Walker, J. E., and Hirst, J. (2001). GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I). J Biol Chem *276*, 38345-38348.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. Embo J *21*, 5843-5852.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat Biotechnol *21*, 89-92.

Forsburg, S. L., and Guarente, L. (1989). Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. Genes Dev *3*, 1166-1178.

Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. Cell *125*, 187-199.

Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. Gene *313*, 17-42.

Fu, H. (2004). Protein-protein interactions : methods and applications (Totowa, N.J.: Humana Press).

Gallagher, J. E., Dunbar, D. A., Granneman, S., Mitchell, B. M., Osheim, Y., Beyer, A. L., and Baserga, S. J. (2004). RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. Genes Dev *18*, 2506-2517.

Gancedo, J. M. (1998). Yeast carbon catabolite repression. Microbiol Mol Biol Rev *62*, 334-361.

Gao, J., Opiteck, G. J., Friedrichs, M. S., Dongre, A. R., and Hefta, S. A. (2003). Changes in the protein expression of yeast as a function of carbon source. J Proteome Res *2*, 643-649.

Gautier, T., Berges, T., Tollervey, D., and Hurt, E. (1997). Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. Mol Cell Biol *17*, 7088-7098.

Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B.*, et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631-636.

Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M.*, et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

Gerbasi, V. R., Weaver, C. M., Hill, S., Friedman, D. B., and Link, A. J. (2004). Yeast Asc1p and mammalian RACK1 are functionally orthologous core 40S ribosomal proteins that repress gene expression. Mol Cell Biol *24*, 8276-8287.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A *100*, 6940-6945.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737-741.

Ghavidel, A., Cagney, G., and Emili, A. (2005). A skeleton of the human protein interactome. Cell *122*, 830-832.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B.*, et al.* (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387-391.

Giorgio Tomasi, F. v. d. B. C. A. (2004). Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. Journal of Chemometrics *18*, 231-241.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E.*, et al.* (2003). A protein interaction map of Drosophila melanogaster. Science *302*, 1727-1736.

Golemis, E., and Adams, P. D. (2005). Protein-protein interactions : a molecular cloning manual, 2nd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Gordon, G. W., Berry, G., Liang, X. H., Levine, B., and Herman, B. (1998). Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. Biophys J *74*, 2702-2713.

Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schafer, T., Kuster, B., Tschochner, H., Tollervey, D.*, et al.* (2002). 90S pre-ribosomes

include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol Cell *10*, 105-115.

Granneman, S., and Baserga, S. J. (2004). Ribosome biogenesis: of knobs and RNA processing. Exp Cell Res *296*, 43-50.

Gu, W., Deng, Y., Zenklusen, D., and Singer, R. H. (2004). A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev *18*, 1452-1465.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol *17*, 994-999.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J.*, et al.* (2004). Transcriptional regulatory code of a eukaryotic genome. Nature *431*, 99-104.

Harnpicharnchai, P., Jakovljevic, J., Horsey, E., Miles, T., Roman, J., Rout, M., Meagher, D., Imai, B., Guo, Y., Brame, C. J.*, et al.* (2001). Composition and functional characterization of yeast 66S ribosome assembly intermediates. Mol Cell *8*, 505-515.

Heinemeyer, W., Trondle, N., Albrecht, G., and Wolf, D. H. (1994). PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits from yeast? Indication for a set of 14 different subunits in the eukaryotic proteasome core. Biochemistry *33*, 12229-12237.

Heitman, J., Movva, N. R., and Hall, M. N. (1991). Targets for cell cycle arrest by the immunosuppressant rapamycin in yeast. Science *253*, 905-909.

Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci U S A *90*, 5011-5015.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981). A gas-liquid solid phase peptide and protein sequenator. J Biol Chem *256*, 7990-7997.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K.*, et al.* (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180-183.

Hong, B., Wu, K., Brockenbrough, J. S., Wu, P., and Aris, J. P. (2001). Temperature sensitive nop2 alleles defective in synthesis of 25S rRNA and large ribosomal subunits in Saccharomyces cerevisiae. Nucleic Acids Res *29*, 2927-2937.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686-691.

Hung, N. J., and Johnson, A. W. (2006). Nuclear recycling of the pre-60S ribosomal subunit-associated factor Arx1 depends on Rei1 in Saccharomyces cerevisiae. Mol Cell Biol *26*, 3718-3727.

Hurt, E., Hannus, S., Schmelzl, B., Lau, D., Tollervey, D., and Simos, G. (1999). A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. J Cell Biol *144*, 389-401.

Iost, I., Dreyfus, M., and Linder, P. (1999). Ded1p, a DEAD-box Protein Required for Translation Initiation in Saccharomyces cerevisiae, Is an RNA Helicase. J Biol Chem *274*, 17677-17683.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics *4*, 1265-1272.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A *98*, 4569-4574.

Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature *409*, 533-538.

Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. Genome Res *12*, 37-46.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., Snyder, M., Greenblatt, J. F., and Gerstein, M. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. Science *302*, 449-453.

Johnson, A. W., Lund, E., and Dahlberg, J. (2002). Nuclear export of ribosomal subunits. Trends in Biochemical Sciences *27*, 580-585.

Jones, A. L., Quimby, B. B., Hood, J. K., Ferrigno, P., Keshava, P. H., Silver, P. A., and Corbett, A. H. (2000). SAC3 may link nuclear protein export to cell cycle progression. Proc Natl Acad Sci U S A *97*, 3224-3229.

Jones, R. B., Gordus, A., Krall, J. A., and MacBeath, G. (2006). A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature *439*, 168-174.

Jorgensen, P., Nishikawa, J. L., Breitkreutz, B. J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science *297*, 395-400.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301.

Kasten, M. M., Dorland, S., and Stillman, D. J. (1997). A large protein complex containing the yeast Sin3p and Rpd3p transcriptional regulators. Mol Cell Biol *17*, 4852-4858.

Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem *74*, 5383-5392.

Killion, P. J., Sherlock, G., and Iyer, V. R. (2003). The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford Microarray Database (SMD). BMC Bioinformatics *4*, 32.

Kiss-Laszlo, Z., Henry, Y., Bachellerie, J. P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell *85*, 1077-1088.

Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W., and Liu, X. (2003). Identification of novel protein-protein interactions using a versatile mammalian tandem affinity purification expression system. Mol Cell Proteomics *2*, 1225-1233.

Koch, C., Moll, T., Neuberg, M., Ahorn, H., and Nasmyth, K. (1993). A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. Science *261*, 1551-1557.

Kressler, D., Linder, P., and de La Cruz, J. (1999). Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. Mol Cell Biol *19*, 7897-7912.

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P.*, et al.* (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637-643.

Krogan, N. J., Peng, W. T., Cagney, G., Robinson, M. D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D. P.*, et al.* (2004). High-definition macromolecular composition of yeast RNA-processing complexes. Mol Cell *13*, 225-239.

Kufel, J., Allmang, C., Petfalski, E., Beggs, J., and Tollervey, D. (2003). Lsm Proteins are required for normal processing and stability of ribosomal RNAs. J Biol Chem *278*, 2147-2156.

Lai, J. S., and Herr, W. (1992). Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. Proc Natl Acad Sci U S A *89*, 6958-6962.

Lamping, E., Luckl, J., Paltauf, F., Henry, S. A., and Kohlwein, S. D. (1994). Isolation and characterization of a mutant of Saccharomyces cerevisiae with pleiotropic deficiencies in transcriptional activation and repression. Genetics *137*, 55-65.

Lebaron, S., Froment, C., Fromont-Racine, M., Rain, J. C., Monsarrat, B., Caizergues-Ferrer, M., and Henry, Y. (2005). The splicing ATPase prp43p is a component of multiple preribosomal particles. Mol Cell Biol *25*, 9269-9282.

Lebreton, A., Saveanu, C., Decourty, L., Rain, J. C., Jacquier, A., and Fromont-Racine, M. (2006). A functional network involved in the recycling of nucleocytoplasmic pre-60S factors. J Cell Biol *173*, 349-360.

Lee, I., Date, S. V., Adai, A. T., and Marcotte, E. M. (2004). A probabilistic functional network of yeast genes. Science *306*, 1555-1558.

Lee, I., Li, Z., and Marcotte, E. M. (2007). An Improved, Bias-Reduced Probabilistic Functional Gene Network of Baker's Yeast, Saccharomyces cerevisiae. PLoS ONE *2*, e988.

Lee, J. H., Pestova, T. V., Shin, B. S., Cao, C., Choi, S. K., and Dever, T. E. (2002). Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. Proc Natl Acad Sci U S A *99*, 16689-16694.

Lee, M. S., Henry, M., and Silver, P. A. (1996). A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev *10*, 1233-1246.

Lee, T. I., and Young, R. A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet *34*, 77-137.

Leeds, N. B., Small, E. C., Hiley, S. L., Hughes, T. R., and Staley, J. P. (2006). The splicing factor Prp43p, a DEAH box ATPase, functions in ribosome biogenesis. Mol Cell Biol *26*, 513-522.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A. G. (2006). Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet *38*, 896-903.

Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T.*, et al.* (2004). A map of the interactome network of the metazoan C. elegans. Science *303*, 540-543.

Lin, H., Nguyen, P., and Vancura, A. (2002). Phospholipase C interacts with Sgd1p and is required for expression of GPD1 and osmoresistance in Saccharomyces cerevisiae. Mol Genet Genomics *267*, 313-320.

Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999). Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol *17*, 676-682.

Liu, H., Sadygov, R. G., and Yates, J. R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem *76*, 4193-4201.

Loewith, R., Jacinto, E., Wullschleger, S., Lorberg, A., Crespo, J. L., Bonenfant, D., Oppliger, W., Jenoe, P., and Hall, M. N. (2002). Two TOR complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control. Mol Cell *10*, 457-468.

Loewith, R., Smith, J. S., Meijer, M., Williams, T. J., Bachman, N., Boeke, J. D., and Young, D. (2001). Pho23 is associated with the Rpd3 histone deacetylase and is required for its normal function in regulation of gene expression and silencing in Saccharomyces cerevisiae. J Biol Chem *276*, 24068-24074.

Longtine, M. S., McKenzie, A., 3rd, Demarini, D. J., Shah, N. G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J. R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953-961.

Lorberg, A., and Hall, M. N. (2004). TOR: the first 10 years. Curr Top Microbiol Immunol *279*, 1-18.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol *25*, 117-124.

MacBeath, G., and Schreiber, S. L. (2000). Printing proteins as microarrays for high-throughput function determination. Science *289*, 1760-1763.

Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999a). Detecting protein function and protein-protein interactions from genome sequences. Science *285*, 751-753.

Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (1999b). A combined algorithm for genome-wide prediction of protein function. Nature *402*, 83-86.

McNabb, D. S., Xing, Y., and Guarente, L. (1995). Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. Genes Dev *9*, 47-58.

Michnick, S. W., Remy, I., Campbell-Valois, F. X., Vallee-Belisle, A., and Pelletier, J. N. (2000). Detection of protein-protein interactions by protein fragment complementation strategies. Methods Enzymol *328*, 208-230.

Milkereit, P., Gadal, O., Podtelejnikov, A., Trumtel, S., Gas, N., Petfalski, E., Tollervey, D., Mann, M., Hurt, E., and Tschochner, H. (2001). Maturation and Intranuclear Transport of Pre-Ribosomes Requires Noc Proteins. Cell *105*, 499-509.

Mitra, R. D., Silva, C. M., and Youvan, D. C. (1996). Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. Gene *173*, 13-17.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., and Tsien, R. Y. (1997). Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature *388*, 882-887.

Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W. T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A.*, et al.* (2004). Exploration of essential gene functions via titratable promoter alleles. Cell *118*, 31-44.

Moy, T. I., and Silver, P. A. (2002). Requirements for the nuclear export of the small ribosomal subunit. J Cell Sci *115*, 2985-2995.

Myer, V. E., and Young, R. A. (1998). RNA polymerase II holoenzymes and subcomplexes. J Biol Chem *273*, 27757-27760.

Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem *75*, 4646-4658.

Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., and Mann, M. (1997). Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc Natl Acad Sci U S A *94*, 385-390.

Ni, L., and Snyder, M. (2001). A genomic study of the bipolar bud site selection pattern in Saccharomyces cerevisiae. Mol Biol Cell *12*, 2147-2170.

Nissan, T. A., Bassler, J., Petfalski, E., Tollervey, D., and Hurt, E. (2002). 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. Embo J *21*, 5539-5547.

Oki, M., and Nishimoto, T. (1998). A protein required for nuclear-protein import, Mog1p, directly interacts with GTP-Gsp1p, the Saccharomyces cerevisiae ran homologue. Proc Natl Acad Sci U S A *95*, 15388-15393.

Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., Resing, K. A., and Ahn, N. G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics *4*, 1487-1502.

Olesen, J., Hahn, S., and Guarente, L. (1987). Yeast HAP2 and HAP3 activators both bind to the CYC1 upstream activation site, UAS2, in an interdependent manner. Cell *51*, 953-961.

Oliver, S. G. (2002). Functional genomics: lessons from yeast. Philos Trans R Soc Lond B Biol Sci *357*, 17-23.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics *1*, 376-386.

Osherovich, L. Z., and Weissman, J. S. (2001). Multiple Gln/Asn-rich prion domains confer susceptibility to induction of the yeast [PSI(+)] prion. Cell *106*, 183-194.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., and Maltsev, N. (1999). The use of gene clusters to infer functional coupling. Proc Natl Acad Sci U S A *96*, 2896-2901.

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci U S A *96*, 4285-4288.

Peng, W.-T., Robinson, M. D., Mnaimneh, S., Krogan, N. J., Cagney, G., Morris, Q., Davierwala, A. P., Grigull, J., Yang, X., Zhang, W.*, et al.* (2003). A Panoramic View of Yeast Noncoding RNA Processing. Cell *113*, 919-933.

Pertschy, B., Saveanu, C., Zisser, G., Lebreton, A., Tengg, M., Jacquier, A., Liebminger, E., Nobis, B., Kappel, L., van der Klei, I.*, et al.* (2007). Cytoplasmic Recycling of 60S Preribosomal Factors Depends on the AAA Protein Drg1. Mol Cell Biol *27*, 6581-6592.

Peterson, C. L., and Tamkun, J. W. (1995). The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem Sci *20*, 143-146.

Piekna-Przybylska, D., Decatur, W. A., and Fournier, M. J. (2007). New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. Rna *13*, 305-312.

Pinkham, J. L., and Guarente, L. (1985). Cloning and molecular analysis of the HAP2 locus: a global regulator of respiratory genes in Saccharomyces cerevisiae. Mol Cell Biol *5*, 3410-3416.

Piper, P. W., Bellatin, J. A., and Lockheart, A. (1983). Altered maturation of sequences at the 3' terminus of 5S gene transcripts in a Saccharomyces cerevisiae mutant that lacks a RNA processing endonuclease. Embo J *2*, 353-359.

Planta, R. J., and Mager, W. H. (1998). The list of cytoplasmic ribosomal proteins of Saccharomyces cerevisiae. Yeast *14*, 471-477.

Reed, R. (1990). Protein composition of mammalian spliceosomes assembled in vitro. Proc Natl Acad Sci U S A *87*, 8031-8035.

Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B. J., Hon, G. C., Myers, C. L., Parsons, A., Friesen, H., Oughtred, R., Tong, A.*, et al.* (2006). Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. J Biol *5*, 11.

Reinke, A., Anderson, S., McCaffery, J. M., Yates, J., 3rd, Aronova, S., Chu, S., Fairclough, S., Iverson, C., Wedaman, K. P., and Powers, T. (2004). TOR complex 1 includes a novel component, Tco89p (YPL180w), and cooperates with Ssd1p to maintain cellular integrity in Saccharomyces cerevisiae. J Biol Chem *279*, 14752-14762.

Rempola, B., Karkusiewicz, I., Piekarska, I., and Rytka, J. (2006). Fcf1p and Fcf2p are novel nucleolar Saccharomyces cerevisiae proteins involved in pre-rRNA processing. Biochem Biophys Res Commun *346*, 546-554.

Remy, I., and Michnick, S. W. (1999). Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. Proc Natl Acad Sci U S A *96*, 5394-5399.

Riego, L., Avendano, A., DeLuna, A., Rodriguez, E., and Gonzalez, A. (2002). GDH1 expression is regulated by GLN3, GCN4, and HAP4 under respiratory growth. Biochem Biophys Res Commun *293*, 79-85.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol *17*, 1030-1032.

Rivas, S., Romeis, T., and Jones, J. D. (2002). The Cf-9 disease resistance protein is present in an approximately 420-kilodalton heteromultimeric membrane-associated complex at one molecule per complex. Plant Cell *14*, 689-702.

Robinson, M. D., Grigull, J., Mohammad, N., and Hughes, T. R. (2002). FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics *3*, 35.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S.*, et al.* (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154-1169.

Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N.*, et al.* (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173-1178.

Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M., and Grunstein, M. (1996). HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. Proc Natl Acad Sci U S A *93*, 14503-14508.

Saleh, A., Collart, M., Martens, J. A., Genereaux, J., Allard, S., Cote, J., and Brandl, C. J. (1998). TOM1p, a yeast hect-domain protein which mediates transcriptional regulation through the ADA/SAGA coactivator complexes. J Mol Biol *282*, 933-946.

Santos-Rosa, H., Moreno, H., Simos, G., Segref, A., Fahrenkrog, B., Pante, N., and Hurt, E. (1998). Nuclear mRNA Export Requires Complex Formation between Mex67p and Mtr2p at the Nuclear Pores. Mol Cell Biol *18*, 6826-6838.

Schafer, T., Strauss, D., Petfalski, E., Tollervey, D., and Hurt, E. (2003). The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. Embo J *22*, 1370-1380.

Schagger, H., and Pfeiffer, K. (2000). Supercomplexes in the respiratory chains of yeast and mammalian mitochondria. Embo J *19*, 1777-1783.

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics *5*, 4-15.

Schmidt, M. C., Kao, C. C., Pei, R., and Berk, A. J. (1989). Yeast TATA-box transcription factor gene. Proc Natl Acad Sci U S A *86*, 7785-7789.

Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. Mol Syst Biol *3*, 88.

Shoemaker, B. A., and Panchenko, A. R. (2007a). Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol *3*, e42.

Shoemaker, B. A., and Panchenko, A. R. (2007b). Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol *3*, e43.

Stevens, S. W., and Abelson, J. (1999). Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. Proc Natl Acad Sci U S A *96*, 7226-7231.

Stryer, L. (1978). Fluorescence energy transfer as a spectroscopic ruler. Annu Rev Biochem *47*, 819-846.

Stryer, L., and Haugland, R. P. (1967). Energy transfer: a spectroscopic ruler. Proc Natl Acad Sci U S A *58*, 719-726.

Taura, T., Krebber, H., and Silver, P. A. (1998). A member of the Ran-binding protein family, Yrb2p, is involved in nuclear protein export. Proc Natl Acad Sci U S A *95*, 7427-7432.

Tharun, S., He, W., Mayes, A. E., Lennertz, P., Beggs, J. D., and Parker, R. (2000). Yeast Sm-like proteins function in mRNA decapping and decay. Nature *404*, 515-518.

Ting, A. Y., Kain, K. H., Klemke, R. L., and Tsien, R. Y. (2001). Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. Proc Natl Acad Sci U S A *98*, 15003-15008.

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H*., et al.* (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364-2368.

Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M*., et al.* (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808-813.

Trapman, J., Retel, J., and Planta, R. J. (1975). Ribosomal precursor particles from yeast. Exp Cell Res *90*, 95-104.

Tschochner, H., and Hurt, E. (2003). Pre-ribosomes on the road from the nucleolus to the cytoplasm. Trends Cell Biol *13*, 255-263.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P.*, et al.* (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623-627.

Utsugi, T., Hirata, A., Sekiguchi, Y., Sasaki, T., Toh-e, A., and Kikuchi, Y. (1999). Yeast tom1 mutant exhibits pleiotropic defects in nuclear division, maintenance of nuclear structure and nucleocytoplasmic transport at high temperatures. Gene *234*, 285-295.

Valasek, L., Mathew, A. A., Shin, B.-S., Nielsen, K. H., Szamecz, B., and Hinnebusch, A. G. (2003). The yeast eIF3 subunits TIF32/a, NIP1/c, and eIF5 make critical connections with the 40S ribosome in vivo. Genes Dev *17*, 786-799.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. Embo J *19*, 1357-1365.

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. Annu Rev Genet *33*, 261-311.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. Nature *417*, 399-403.

Wade, C. H., Umbarger, M. A., and McAlear, M. A. (2006). The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. Yeast *23*, 293-306.

Walker, M. G., Volkmuth, W., Sprinzak, E., Hodgson, D., and Klingler, T. (1999). Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. Genome Res *9*, 1198-1203.

Wang, W., Zhou, H., Lin, H., Roy, S., Shaler, T. A., Hill, L. R., Norton, S., Kumar, P., Anderle, M., and Becker, C. H. (2003). Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem *75*, 4818-4826.

Warner, J. R., Knopf, P. M., and Rich, A. (1963). A multiple ribosomal structure in protein synthesis. Proc Natl Acad Sci U S A *49*, 122-129.

Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol *19*, 242-247.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H.*, et al.* (1999). Functional

characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science *285*, 901-906.

Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem *73*, 5683-5690.

Wouters, F. S., and Bastiaens, P. I. (1999). Fluorescence lifetime imaging of receptor tyrosine kinase activity in cells. Curr Biol *9*, 1127-1130.

Yao, W., Roser, D., Kohler, A., Bradatsch, B., Bassler, J., and Hurt, E. (2007). Nuclear export of ribosomal 60S subunits by the general mRNA export receptor Mex67-Mtr2. Mol Cell *26*, 51-62.

Young, R. A. (1991). RNA polymerase II. Annu Rev Biochem *60*, 689-715.

Young, R. A., and Davis, R. W. (1983). Yeast RNA polymerase II genes: isolation with antibody probes. Science *222*, 778-782.

Yu, X., Xu, D., and Cheng, Q. (2006). Label-free detection methods for protein microarrays. Proteomics *6*, 5493-5503.

Zemp, I., and Kutay, U. (2007). Nuclear export and cytoplasmic maturation of ribosomal subunits. FEBS Lett *581*, 2783-2793.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T.*, et al.* (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101-2105.

Vita

Zhihua Li graduated from Shijiazhuang First High School, Hebei province, China in 1991. Then he entered Nankai University, Tianjin, China and received the degree of Bachelor of Science in Biochemistry in 1995. He joined the Institute of Genetics, Chinese Academy of Sciences, Beijing, China as a graduate student and received the degree of Master of Science in Genetics in 1998. He then went to the University of Southern California and received the degree of Master of Science in Microbiology and Immunology in 2000. He continued his graduate study in the Cell and Molecular Biology Program in the University of Texas at Austin since then. While working toward his Ph.D., he also received a degree of Master of Science in Statistics from the University of Texas at Austin in 2006.

Permanent address: Shijiazhuang, Hebei Province, China 050031

This dissertation was typed by the author.